ABSTRACT

Title of dissertation: TOPICS IN EQUILIBRIUM AND NONEQUILIBRIUM THERMODYNAMICS: COMPUTING CRYSTALLINE FREE ENERGIES AND ENGINEERING MAXWELL'S DEMON

Zhiyue Lu, Doctor of Philosophy, 2015

Dissertation directed by: Professor Christopher Jarzynski
Department of Chemistry and Biochemistry;
Institute for Physical Science and Technology

This dissertation covers two separate topics in statistical physics.

The first part of the dissertation focuses on computational methods of obtaining the free energies (or partition functions) of crystalline solids. We describe a method to compute the Helmholtz free energy of a crystalline solid by direct evaluation of the partition function. In the many-dimensional conformation space of all possible arrangements of N particles inside a periodic box, the energy landscape consists of localized islands corresponding to different solid phases. Calculating the partition function for a specific phase involves integrating over the corresponding island. Introducing a natural order parameter that quantifies the net displacement of particles from lattices sites, we write the partition function in terms of a one-dimensional integral along the order parameter, and evaluate this integral using umbrella sampling. We validate the method by computing free energies of both

face-centered cubic (FCC) and hexagonal close-packed (HCP) hard sphere crystals with a precision of $10^{-5}k_BT$ per particle. In developing the numerical method, we find several scaling properties of crystalline solids in the thermodynamic limit. Using these scaling properties, we derive an explicit asymptotic formula for the free energy per particle in the thermodynamic limit. In addition, we describe several changes of coordinates that can be used to separate internal degrees of freedom from external, translational degrees of freedom.

The second part of the dissertation focuses on engineering idealized physical devices that work as Maxwell's demon. We describe two autonomous mechanical devices that extract energy from a single heat bath and convert it into work, while writing information onto memory registers. Additionally, both devices can operate as Landauer's eraser, namely they can erase information from a memory register, while energy is dissipated into the heat bath. The phase diagrams and the efficiencies of the two models are solved and analyzed. These two models provide concrete physical illustrations of the thermodynamic consequences of information processing.

TOPICS IN EQUILIBRIUM AND NONEQUILIBRIUM THERMODYNAMICS: COMPUTING CRYSTALLINE FREE ENERGIES AND ENGINEERING MAXWELL'S DEMON

by

Zhiyue Lu

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2015

Advisory Committee:
Professor Christopher Jarzynski, Chair/Advisor
Professor John D. Weeks
Professor Garegin A. Papoian
Professor Victor M. Yakovenko
Professor Perinkulam S. Krishnaprasad, Dean's Representative

Acknowledgments

Many people have deeply influenced me. They may or may not know how important the role they have played in my life. I would like to thank them all for their guidance, friendship, wisdom, and abundant love and support.

First of all, I would like to thank my advisor and mentor, Prof. Christopher Jarzynski. Chris is my role model with extraordinary scientific clarity and clear logic. I enjoy each discussion I had with him, since he can use precise language and explicit examples to explain complicated theory in such a clear and understandable way. In the past 5 years, through his patient guidance, Chris has taught me how to think logically and precisely. Thank you Chris, for providing me with these important tools for becoming a scientist. In addition, thank you for being my PhD advisor and thank you for the valuable lessons and memories. I can't make it here without your abundant care and support.

Thank you Prof. Garegin Papoian, for bringing me from North Carolina to Maryland. Without you, my journey as a graduate student can't even start. Thank you, Prof. Michael Coplan for taking care of we graduate students in Chemical Physics Program and spending your time with us. Thank you, Prof. James Yorke for your wisdom and guidance in the summer of 2011. Thank you Prof. John Weeks for the inspiring discussions on hard sphere crystals. I would like to thank all of my PhD exam committee members for their suggestions to the dissertation and the inspiring questions they have asked during and after the PhD exam.

I would like to thank my friends and colleagues for the inspiring discussions

and sweet friendship. Thank you Dr. Dibyendu Mandal, Dr. Andy Ballard, Dr. Shaon Chakrabarti, Dr. Sebastian Deffner, Dr. Haitao Quan, Dr. Fei Liu, Dr. Oren Raz, Dr. Yigit Subasi, Jeffrey Demers, Ayoti Patra, Andrew Smith, Dr. Lifang Ma, Wentao Song, Yi-Hsieh Wang, Dr. Mick Warehime and Debbie Jenkins. You made my graduate school life full of joy and love.

Thank you Zhixin Lu, for giving me so much joy and for sharing the privilege of having a twin brother. Many adventures can't happen without you. Thank you Mom and Dad for your love. I know sometimes I make you worry, but thanks for giving me the freedom to explore. I also thank my wife Dr. Huizhi Bai, for the road we have traveled together, and for the life we have shared with each other. Thank you Dr. Theodore Pickett, Jr. for providing me with the wisdom and the tools to face my fear. I will never quit learning and growing. I would also like to thank all of my friends from Maryland Chinese Bible Study Group for their abundant support and love.

Thank you Lord, for the creation and for revealing your wisdom. Through your amazing grace I learn to be humble.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

FEP	Free energy perturbation
TI	Thermodynamic integration
FCC	Face centered cubic
HCP	Hexagonal close packed
FL	Frenkel and Ladd method
LS	Lattice switch method
MC	Monte Carlo,
MCMC	Markov chain Monte Carlo
WHAM	Weighted histogram analysis method
B.C.	Boundary conditions
r.m.s.	Root mean square
MJ	Mandal and Jarzynski's model
CW	Clockwise
CCW	Counterclockwise

Chapter 1: Overview of Computational Free Energies

1.1 Background

Given a classical macroscopic system in equilibrium with a heat bath at temperature T, and with both its volume V and particle number N fixed, each particle evolves erratically. Instead of getting lost in studying the dynamical trajectory of the enormous number of particles ($N \approx 10^{20}$), it is more convenient to study the collective properties of the whole system. In thermodynamics, state functions are used to characterize systems' collective properties. Among the many state functions (e.g. entropy S, internal energy U, pressure P, chemical potential μ), the Helmholtz free energy F describes both the thermodynamic stability and the amount of work that can be extracted isothermally from a system of interest. The Helmholtz free energy is defined in thermodynamics as [4]

$$F = U - TS \tag{1.1}$$

Free energies are frequently used in studying phase equilibria, determining spontaneous directions of chemical reactions, comparing relative stabilities of metastable states (e.g. different conformations of folded proteins in biological systems), etc.

In addition, the free energy measures the maximum work that one can extract by isothermally transforming a system from one state (A) to another (B). This is formally stated in the principle of maximum work:

$$W \leq T\Delta S - \Delta U = -\Delta F = F_A - F_B \tag{1.2}$$

where W denotes the work delivered by the system, $\Delta S = S_B - S_A$ denotes the difference of entropies between states A and B, $\Delta U = S_B - S_A$ denotes the difference of internal energies between A and B, and $\Delta F = F_B - F_A$ denotes the free energy difference between A and B. Note that the equality is achieved when the transformation is carried out reversibly and isothermally, in other words, the system is in equilibrium with the surroundings at each instant in time.

In classical statistical physics, thermal equilibrium is described from an ensemble point of view [5]. A system of interest can be considered as an ensemble of microscopic states (microstates) that follows a certain probability distribution in the phase space. Each microstate fully specifies the positions and momenta of all the particles of the system. Specifically, for an ensemble of microstates at fixed N, V, T (or a canonical ensemble), the equilibrium probability distribution of the microstates is proportional to Boltzmann's factor, normalized by the partition function Z. In other words, Z is simply a weighted measure of the phase space with the weighting factor given by Boltzmann's factor $e^{-\beta H}$,

$$Z = \frac{1}{N! h^{dN}} \int \mathrm{d}\mathbf{\Gamma}\ e^{-\beta H(\mathbf{\Gamma})} \tag{1.3}$$

where h is the Planck constant, $\beta^{-1} = k_B T$ is Boltzmann's constant times the temperature, $H(\mathbf{\Gamma})$ is the Hamiltonian of the system of interest and $\mathbf{\Gamma} = (\mathbf{q}, \mathbf{p}) \in \mathbb{R}^{DN}$ denotes a microstate. Here for simplicity we assume the N particles located in D-dimensional space are indistinguishable. The Helmholtz free energy F can be computed from the value of the partition function [6] by:

$$F = -\beta^{-1} \ln Z \tag{1.4}$$

In addition, we can write the free energy difference between two systems as

$$\Delta F = F_B - F_A = \beta^{-1} \ln \frac{Z_A}{Z_B} \tag{1.5}$$

where Z_A and Z_B are the canonical partition functions of systems A and B.

Although Eq. 1.3 and Eq. 1.4 are simple formulas that define the free energy, in practice it is very difficult to compute the free energy by directly evaluating the multi-dimensional integral over the phase space (see Eq. 1.3). To date, most numerical methods [7–23] have been developed for computing free energy differences, instead of the absolute values of free energies. In this chapter, I briefly review some numerical methods for computing free energy differences of crystalline solids. In Chapter 2, I describe a new numerical method for computing the absolute free energies of crystalline solids. In this new method, crystalline free energies are computed by direct evaluations of the canonical partition functions. Directly evaluating the absolute free energies allows one to compare free energy differences between two

distinct systems to which the other methods are difficult to apply. Inspired by the work described in Chapter 2, in Chapter 3 I derive an asymptotic formula for the free energy per particle for crystalline solids in the thermodynamic limit. In addition, we found in Chapter 2 that for numerical simulations of systems with periodic boundary conditions, it is important to separate the translational degrees of freedom of the entire system from the internal degrees of freedom. In Chapter 4, I present and describe several changes of coordinates that separate the translational and the internal degrees of freedom. These changes of coordinates are potentially useful in evaluating free energies of systems simulated with periodic boundary conditions.

1.2 Computing free energy differences – general methods

In statistical physics, there are several general frameworks for computing relative free energies. Without losing generality, let us assume $H(\mathbf{\Gamma})$, the Hamiltonian of the system of interest consists of two terms, the kinetic energy and potential energy, depending respectively on $\mathbf{p}$ and $\mathbf{q}$,

$$H = \frac{1}{2m} |\mathbf{p}|^2 + V(\mathbf{q}) \quad . \tag{1.6}$$

When the Hamiltonian is in this form, the partition function can be factorized into two independent integrals, and the integral over the momenta can be analytically calculated by evaluating products of independent Gaussian integrals. We write the

partition function as follows:

$$Z = \frac{1}{N! h^{3N}} \int \mathrm{d}\mathbf{\Gamma}\ e^{-\beta H(\mathbf{\Gamma})} = \frac{(2\pi m \beta^{-1})^{3N/2}}{h^{3N} N!} \cdot Q \quad , \tag{1.7}$$

where Q denotes the configuration integral:

$$Q = \int \mathrm{d}\mathbf{q}\ e^{-\beta V(\mathbf{q})} \quad . \tag{1.8}$$

Although in this chapter we do not specify the domain of the integral in Eq. 1.8, one should consider the domain as a bounded region in $\mathbb{R}^{DN}$ whose boundary is influenced by the geometrical boundary (container) of the system. In fact, it is a non-trivial problem to correctly define this domain, especially when the system has multiple regions of configuration space corresponding to different phases. A discussion of this topic can be found in the next chapter.

1.2.1 Free energy perturbation method (FEP)

The free energy perturbation method, based on an equality derived by Zwanzig [24], provides the free energy difference between two systems as

$$\begin{aligned} \Delta F &= -\beta^{-1} \ln \left(\frac{Z_1}{Z_0} \right) \\ &= -\beta^{-1} \ln \left(\frac{\int \mathrm{d}\mathbf{q}\ e^{-\beta V_1}}{\int \mathrm{d}\mathbf{q}\ e^{-\beta V_0}} \right) \\ &= -\beta^{-1} \ln \left(\frac{\int \mathrm{d}\mathbf{q}\ e^{-\beta V_0} \cdot e^{-\beta \Delta V}}{\int \mathrm{d}\mathbf{q}\ e^{-\beta V_0}} \right) \\ &= -\beta^{-1} \ln \left\langle e^{-\beta \Delta V} \right\rangle_0 \quad . \end{aligned} \tag{1.9}$$

Here, the angular brackets $\langle\cdot\rangle_0$ denote an average over the equilibrium distribution corresponding to the potential $V_0(\mathbf{q})$, and $\Delta V(\mathbf{q}) = V_1(\mathbf{q}) - V_0(\mathbf{q})$. Eq. 1.9 reveals that the free energy difference ΔF can be calculated from the equilibrium average of the quantity $e^{-\beta\Delta V(\mathbf{q})}$. Note that some configurations contribute significantly to Z_1 and some configurations contributes significantly to Z_2, we refer to these as dominant domains. If there is little overlap between the dominant domains corresponding to Z_1 and Z_2, then the average $\left\langle e^{-\beta\Delta V}\right\rangle_0$ will converge slowly. In this case, more advanced sampling schemes become necessary. For further reading, we refer the reader to Ref. [25–28].

1.2.2 Thermodynamic integration method (TI)

The thermodynamic integration method is based on an equality introduced by Kirkwood [29]. Consider two potential functions, $V(\mathbf{q};\lambda = 0)$ and $V(\mathbf{q};\lambda = 1)$ corresponding to two systems 0 and 1. The parameter λ is externally controlled. As λ is varied parametrically from 0 to 1, the potential function $V(\mathbf{q};\lambda)$ changes smoothly from system 0 to system 1. Along the path, we have

$$\begin{aligned}\frac{\partial F(\lambda)}{\partial\lambda} &= -\beta^{-1}\frac{\partial \ln Z_\lambda}{\partial\lambda} \\ &= -\beta^{-1}\left[\int \mathrm{d}\mathbf{q}(-\beta)e^{-\beta V}\frac{\partial V}{\partial\lambda}\right]\Big/ Z_\lambda \\ &= \left\langle\frac{\partial V(\mathbf{q};\,\lambda)}{\partial\lambda}\right\rangle_\lambda \quad .\end{aligned} \tag{1.10}$$

By integrating both sides of Eq. 1.10, one can get the free energy difference:

$$\begin{aligned}\Delta F &= \int_0^1 \mathrm{d}\lambda \, \frac{\partial F(\lambda)}{\partial \lambda} = \int_0^1 \mathrm{d}\lambda \, \left\langle \frac{\partial V(\mathbf{q};\, \lambda)}{\partial \lambda} \right\rangle_\lambda \\ &\approx \sum_{i=1}^{n-1} \left\langle \frac{\partial V}{\partial \lambda}(\mathbf{q};\, \lambda_i) \right\rangle_{\lambda_i} \delta\lambda_i\end{aligned} \tag{1.11}$$

where $\delta\lambda_i = \lambda_{i+1} - \lambda_i$. In practice, the integral is approximately evaluated as a sum, where a set of equilibrium averages of $\partial V/\partial\lambda$ are evaluated at a set of λ_i's starting at $\lambda_1 = 0$ and ending at $\lambda_n = 1$. To ensure accuracy, the $\delta\lambda_i$'s need to be sufficiently small. Additionally, the path $\lambda : 0 \to 1$ should avoid going through phase transitions, such as crystallization or melting, where the value of $\partial F/\partial\lambda$ may become singular.

1.3 Crystalline solids

In Sec. 1.2, we reviewed two important theoretical frameworks for computing free energy differences between different systems. In the following, we narrow our scope to solids.

The word "solid" can denote a crystalline solid, which is periodic in space. It can also denote amorphous but mechanically stable materials, where particles are randomly packed into a frustrated configuration. In contrast to a liquid or gas, a solid has a mechanically stable structure. For crystalline solid, one can define a set of perfect lattice sites that are periodic in space. When the center of mass of the entire crystal is fixed, each particle fluctuates erratically around its site (see Fig. 2.1), without diffusing away as would be the case with a liquid or gas. To set

up notations, let us denote the position of the i^{th} particle in 3-dimensional space by $\vec{q}_i \in \mathbb{R}^3$. The particle's *reference position* is the position of its perfect lattice site, denoted by $\vec{q}_i^{\,\mathrm{lat}} \in \mathbb{R}^3$. With a well defined set of reference positions, one can describe particle positions either by $\vec{q}_i$ or the displacement vector $\vec{r}_i = \vec{q}_i - \vec{q}_i^{\,\mathrm{lat}}$. This dual description of the positions of particles can be generalized to a dual description of the configuration of the crystal:

$$\mathbf{r} = \mathbf{q} - \mathbf{q}^{\mathrm{lat}} = (\vec{r}_1,\ \vec{r}_2,\ \cdots \vec{r}_N) \in \mathbb{R}^{3N} \quad . \tag{1.12}$$

In the **r**-space description, the origin $\mathbf{r} = \mathbf{0}$ represents the reference configuration where every particle sits exactly at its perfect lattice site (or reference position).

1.3.1 Two simple models of solids

There are two simple models of crystalline solids – the Einstein crystal and the hard sphere crystal.

An Einstein crystal is essentially a collection of non-interacting harmonic oscillators. In the Einstein crystal, each particle vibrates independently under a harmonic potential associated with its own lattice site, and no interaction occurs between particles. Thus, the potential of the Einstein crystal is simply a sum of harmonic potentials,

$$V^{\mathrm{E}}(\mathbf{q}) = \sum_i \frac{1}{2} k \left(\vec{q}_i - \vec{q}_i^{\,\mathrm{lat}}\right)^2 = \frac{1}{2} k \left|\mathbf{r}\right|^2 \quad . \tag{1.13}$$

Although this system is overly simplified, in particular there are no particle-particle interactions, it is a very convenient model of a crystalline solid since its free energy can be calculated exactly by evaluating Gaussian integrals

$$Q^{\mathrm{E}} = \int \mathrm{d}\mathbf{r}\ e^{-\beta k|\mathbf{r}|^2/2} = \left(\frac{2\pi}{\beta k}\right)^{3N/2} \quad . \tag{1.14}$$

There are many other models (e.g. Debye model, Lennard-Jones crystals and so on) of crystalline solids where the particles do interact with each other. Conceptually the simplest model that allows for particle interactions is the hard sphere crystal – a number of hard spheres confined in a finite volume form a stable crystal when the volume density is sufficiently high. This model has been widely studied in numerical simulations.

In a hard sphere crystal, particles interact under the pairwise hard interaction potential,

$$V_{\mathrm{HS}}\left(\vec{q}_i,\ \vec{q}_j\right) = \begin{cases} \infty & ,\ |\vec{q}_i - \vec{q}_j| \leq d \\ 0 & ,\ |\vec{q}_i - \vec{q}_j| > d \end{cases} \quad , \tag{1.15}$$

where d is the diameter of each particle. The total potential energy is

$$V(\mathbf{q}) = \sum_{i<j} V_{\mathrm{HS}}\left(\vec{q}_i,\ \vec{q}_j\right) \quad , \tag{1.16}$$

which is 0 when no pairs of spheres overlap, and is infinite otherwise. In hard sphere crystals, only non-overlapping configurations are energetically *allowed configurations*.

1.3.2 Polytypes of crystalline solids

In hard sphere crystalline solids, when the density is above melting, a single canonical ensemble (N, V, T) may contain configurations that corresponds to multiple crystalline phases. A good example is that both face centered cubic (FCC) and hexagonal close packing (HCP) phases exist even in the limit of a closely packed hard sphere crystal.

What is more, allowable configurations include not only FCC and HCP, but also many other phases (polytypes). To see this, let us consider a hard sphere crystal arranged in stacks of hexagonally packed layers, labeled A, B, and C. They differ from each other by a horizontal translation (or shift). The FCC arrangement corresponds to a periodic packing pattern ABCABC... and HCP corresponds to ABAB... Even at the maximum density (closely packed), the only rule of stacking is that two neighboring layers cannot be of the same type. Thus, both ...ABABCBACBA... and ...ABCABCABAB... are allowed orders of stacking, but they are neither FCC nor HCP.

This thesis focus on FCC and HCP crystals, but it is important to acknowledge the existence of multiple polytypes even in a single canonical ensemble.

1.4 Free energies of crystalline solids

In the following, we first review the Frenkel and Ladd method [7], which is widely used for computing free energies of crystalline solids. We then review the lattice switch method [9], specifically designed for evaluating the free energy difference

between solid phases (e.g., FCC and HCP crystals).

1.4.1 Frenkel and Ladd Method

In 1984, Frenkel and Ladd proposed a Monte Carlo method to compute the absolute Helmholtz free energy of an arbitrary crystalline solid [7]. The basic idea behind the Frenkel and Ladd method (FL) is straightforward. By connecting the crystalline solid described by the potential $V_1(\mathbf{q})$ to a reference crystal described by $V_0(\mathbf{q})$, whose free energy F_0 is known, the absolute free energy of the crystalline solid F_1 is obtained by evaluating the free energy difference ΔF between the crystalline solid and the reference crystal. Usually the thermodynamic integration method is used to calculate ΔF.

As mentioned in Sec. 1.2.2, the choice of the reference crystalline potential $V_0(\mathbf{q})$ and thus the path connecting the crystalline solid and the reference crystal are important. Frenkel and Ladd chose the Einstein crystal as the reference crystal, and the path is described in the following paragraph.

The tunable potential connecting the Einstein crystal and the hard sphere crystalline solid is defined as

$$V(\mathbf{r};\lambda) = V^{\mathrm{HS}}(\mathbf{r}) + \lambda \cdot |\mathbf{r}|^2 \quad . \tag{1.17}$$

When $\lambda = 0$, the tunable potential is simply the hard sphere potential. When λ is very large, the second term in Eq. 1.17 dominates the potential, which becomes very close to that of the Einstein crystal. Let λ_E denote a cutoff value, defined

so that a system described by the potential $V(\mathbf{r};\lambda_E)$ can be approximately treated as an Einstein crystal. The path of $\lambda : \lambda_E \to 0$ smoothly connects the potential of an approximate Einstein crystal to the potential of the hard sphere crystalline solid. To implement the thermodynamic integration method (Eq. 1.11), a series of Monte Carlo simulations are performed along the path, and the ensemble average of $\partial V/\partial\lambda = |\mathbf{r}|^2$ is evaluated at each intermediate $\lambda \in [0, \lambda_E]$.

The choice of the value λ_E influences the accuracy and the efficiency of the free energy evaluation. On the one hand, the larger the choice of λ_E, the more accurate the Einstein crystal approximation. On the other hand, when λ_E is larger, more computational time is consumed in the thermodynamic integration.

1.4.2 Lattice switch method

The lattice switch method [9] was introduced to compute the free energy difference between HCP and FCC hard sphere crystals. In what follows, we review the lattice switch operation within configuration space. Then we illustrate the basic idea of the lattice switch Monte Carlo by comparing it to conventional Monte Carlo.

First, let us recall the description of crystalline solids in Eq. 1.12. With a reference configuration $\mathbf{q}^{\text{lat}}$, the displacement vector $\mathbf{r}$ can be mapped to a crystal configuration $\mathbf{q}$. By varying $\mathbf{q}^{\text{lat}}$, the same displacement vector $\mathbf{r}$ can be mapped to different crystalline structures. Taking reference configurations of FCC and HCP crystals, the same displacement vector $\mathbf{r}$ can be used to describe two configurations: $\mathbf{q}_{\text{HCP}} = \mathbf{r} + \mathbf{q}^{\text{lat}}_{\text{HCP}}$ and $\mathbf{q}_{\text{FCC}} = \mathbf{r} + \mathbf{q}^{\text{lat}}_{\text{FCC}}$. A Monte Carlo trial move in which the

reference configuration is switched between $\mathbf{q}_{\mathrm{HCP}}^{\mathrm{lat}}$ and $\mathbf{q}_{\mathrm{FCC}}^{\mathrm{lat}}$, at fixed $\mathbf{r}$, is called a *lattice switch* operation. Let us define

$$\mathbf{A} = \left\{ \mathbf{r} \middle| \; V^{\mathrm{HS}}\left(\mathbf{r} + \mathbf{q}_{\mathrm{HCP}}^{\mathrm{lat}}\right) = 0 \right\} \tag{1.18}$$

as the set of displacement vectors corresponding to allowed HCP configurations, and

$$\mathbf{B} = \left\{ \mathbf{r} \middle| \; V^{\mathrm{HS}}\left(\mathbf{r} + \mathbf{q}_{\mathrm{FCC}}^{\mathrm{lat}}\right) = 0 \right\} \tag{1.19}$$

as the set corresponding to allowed FCC arrangements. Their intersection $\mathbf{C} = \mathbf{A} \cap \mathbf{B}$, consists of the small fraction of displacement vectors, including $\mathbf{r} = \mathbf{0}$, that correspond to allowed FCC and HCP configurations. These sets are schematically sketched in Fig. 1.1.

A conventional Markov Chain Monte Carlo (MCMC) simulation starting from an HCP configuration generates a chain of trial moves that attempt to change to the configuration, and each trial move is followed by acceptance or rejection based on the Metropolis rule [30]. Due to the huge entropic barrier between the set of allowed HCP configurations and the set of FCC configurations, it is unlikely for a trial move in $\mathbf{q}$-space to successfully transform an HCP configuration to an allowed FCC configuration. In the $\mathbf{q}$-space, a trajectory generated by such a conventional MCMC is sketched as a dashed line in Fig. 1.1. Notice that the trajectory is remain confined to HCP configurations.

Within a lattice switch simulation, the conventional MCMC simulation is com-

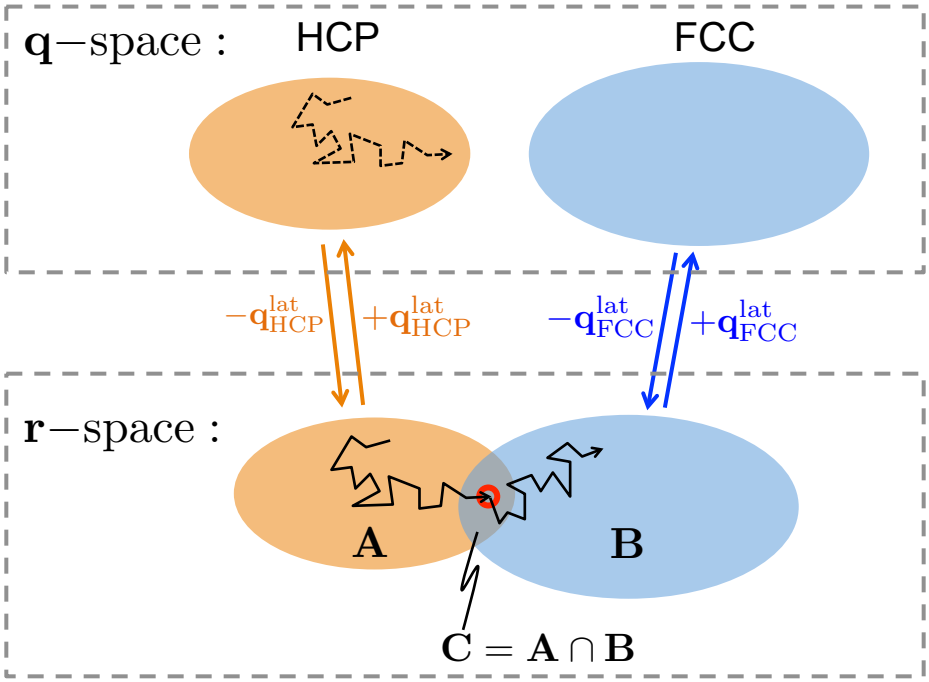


Figure 1.1: The sets of allowed configurations are sketched in **q**-space and **r**-space. Note that the allowed HCP configurations (yellow) and FCC configurations (blue) in **q**-space can be mapped to **A** and **B** in **r**-space respectively by $\mathbf{r} = \mathbf{q} - \mathbf{q}^{\text{lat}}_{\text{HCP}}$ and $\mathbf{r} = \mathbf{q} - \mathbf{q}^{\text{lat}}_{\text{FCC}}$. The dashed line shows a trajectory of configurations sampled by a conventional MCMC simulation in **q**-space. In the LS method, by mapping the MCMC trajectories to **r**-space and introducing lattice switches (shown as the red circle), both FCC and HCP configurations are sampled in one simulation. $\mathbf{C} = \mathbf{A} \cap \mathbf{B}$ is the gateway region where lattice switch operations are accepted by the Metropolis rule. In reality, the measure of **C** is much smaller than the measure of **A** and **B**.

bined with a series of lattice switch operations. First, a conventional MCMC simulation generates a trajectory (in **q**-space) of allowed HCP configurations starting from a random HCP configuration. This trajectory in **q**-space can be mapped onto **r**-space by the following map:

$$\mathbf{r} = \mathbf{q} - \mathbf{q}^{\text{lat}}_{\text{HCP}}. \tag{1.20}$$

Since the trajectory is restricted to HCP configurations in **q**-space, the corresponding trajectory in the **r**-space is confined in **A**. In a LS simulation, after every n steps of MCMC simulation, a lattice switch operations is made, followed by acceptance and rejection based on the Metropolis rule. A lattice switch from HCP to FCC can be written as

$$\text{LS}(\mathbf{q}) = \mathbf{r} + \mathbf{q}^{\text{lat}}_{\text{FCC}} = \mathbf{q} - \mathbf{q}^{\text{lat}}_{\text{HCP}} + \mathbf{q}^{\text{lat}}_{\text{FCC}} \tag{1.21}$$

where **r** is found in Eq. 1.20. If the new configuration LS(**q**) has particle-particle overlap, then it is rejected by the Metropolis rule, and the simulation returns to perform another n steps of MCMC before it make the next attempt of lattice switch. If the new configuration is accepted ($\mathbf{r} \in \mathbf{C}$), the next n steps of MCMC simulation are performed for FCC configurations. The region where the lattice switch is allowed is called the *gateway region*. In Fig. 1.1, an accepted lattice switch is shown as a red circle. After many lattice switch operations, the simulation is able to generate a trajectory that visits back and forth between FCC and HCP configurations, and accumulates samples from both FCC and HCP crystals in a single simulation. As

the number of accepted lattice switches increases, the ratio between the accumulated samples from HCP and FCC crystal approaches the ratio between the measures of **A** and **B**. For hard sphere crystals we can write,

$$\frac{Z^{\mathrm{HCP}}}{Z^{\mathrm{FCC}}} = \frac{\text{the measure of } \mathbf{A}}{\text{the measure of } \mathbf{B}} \quad . \tag{1.22}$$

Finally, we can calculate the free energy difference between HCP and FCC hard sphere crystals

$$F^{\mathrm{FCC}} - F^{\mathrm{HCP}} = -\frac{1}{\beta} \ln \frac{Z^{\mathrm{FCC}}}{Z^{\mathrm{HCP}}} \quad . \tag{1.23}$$

In practice, the accuracy of the free energy difference depends critically on the number of accepted switches between HCP and FCC crystals. Unfortunately, since the measure of gateway region **C** is negligible compared to the measure of **A** or **B**, the frequency of accepted switches is still lower than acceptable in the unbiased Monte Carlo sampling. The reader is referred to the original reference [9], where a flat histogram Monte Carlo simulation method is used to further enhance the transition frequency.

Chapter 2: Integral Method of Computing Free Energies of Crystalline Solids

2.1 Introduction

As discussed in the previous chapter, a variety of computational methods have been developed to evaluate free energies or free energy differences of crystalline solids. The widely used *Einstein crystal* approach, due to Frenkel and Ladd [7, 8], uses thermodynamic integration to compute the free energy difference between a structure of interest and an ideal Einstein crystal. In the *lattice switch* method of Bruce *et al* [9, 10], transitions are generated between two solid phases, whose free energy difference is then obtained from the relative frequencies with which the two phases are sampled. In these and other methods [11–23], the focus is on computing a free energy difference between two structures, one of which might be a reference system whose free energy is known.

Here we describe a method in which the Helmholtz free energy of a crystalline solid is obtained by direct evaluation of the corresponding partition function. Although the method is general, for convenience we describe it in the particular context of the FCC hard sphere crystal.

The derivation of our method involves a few key steps. After defining the problem, we introduce coordinates that enable us to integrate out the three translational degrees of freedom of the entire crystal (Eqs. 2.10-2.18), thereby avoiding the need to add explicit correction terms associated with the center-of-mass constraint [31]. In the remaining degrees of freedom (the conformation space), the energy landscape forms a set of localized islands representing stable structures (Fig. 2.2). The FCC partition function is proportional to the volume of one such island (Eq. 2.22). We then use an order parameter, r, to write this volume as a one-dimensional integral (Eq. 2.23), which we evaluate using standard importance sampling techniques. We apply our method to compute the absolute free energies (per particle) of both FCC and HCP hard sphere crystals. Our results for the difference $f_{\mathrm{HCP}} - f_{\mathrm{FCC}}$ agree with those of the lattice switch method [1].

At the end of this chapter, we use our approach to derive a compact formula for the free energy per particle in the thermodynamic limit (Eq. 3.19).

Our derivation involves technical derivations that are described in detail in Chapter 3, 4 and Appendix, but the application of the method described in this chapter in a numerical simulation is straightforward.

2.2 Theory

2.2.1 Defining the problem

Consider N identical particles of mass m inside a rectangular region, or box, $\mathbf{\Omega}$, of dimensions $l_x \times l_y \times l_z$, with periodic boundary conditions at all six faces. We

denote momenta and positions as

$$\begin{aligned} \mathbf{p} &= (\vec{p}_1,\ \vec{p}_2,\ \cdots,\ \vec{p}_N) \in \mathbb{R}^{3N} \\ \mathbf{q} &= (\vec{q}_1,\ \vec{q}_2,\ \cdots,\ \vec{q}_N) \in \mathbf{\Omega}^N \end{aligned} \tag{2.1}$$

and use $\mathbf{\Gamma} = (\mathbf{p}, \mathbf{q})$ to specify a point in $6N$-dimensional phase space. We take a Hamiltonian of the form

$$H(\mathbf{\Gamma}) = \sum_i \frac{|\vec{p}_i|^2}{2m} + V(\vec{q}_1,\ \vec{q}_2,\ \cdots,\ \vec{q}_N) \tag{2.2}$$

and require the many-particle potential V to be translationally invariant. That is, the value of V is unchanged if every particle undergoes a translation

$$\vec{q}_i \to \left[\vec{q}_i + \vec{\alpha}\right]_{\mathbf{\Omega}} \quad , \quad 1 \leq i \leq N \tag{2.3}$$

for arbitrary $\vec{\alpha} \in \mathbb{R}^3$. Throughout the thesis, the notation $[\vec{q}]_{\mathbf{\Omega}}$ indicates the periodic image of $\vec{q}$ within $\mathbf{\Omega}$, thereby enforcing periodic boundary conditions. The full partition function for this system is

$$Z = \frac{1}{N! h^{3N}} \int \mathrm{d}\mathbf{\Gamma} e^{-\beta H(\mathbf{\Gamma})} = \frac{Q}{\Lambda^{3N} N!} \tag{2.4}$$

where $\beta = 1/k_B T$, $\Lambda = \sqrt{\beta h^2 / 2\pi m}$ is the thermal de Broglie wavelength, and

$$Q = \int_{\mathbf{\Omega}^N} \mathrm{d}\mathbf{q}\, e^{-\beta V(\mathbf{q})}\,. \tag{2.5}$$

In Eq. 2.5 the integration is performed over all possible arrangements of N particles in the box, but typically one is interested in the contribution Q_α from a particular phase, such as the face centered cubic (α = FCC) or hexagonal close packed (α = HCP) phase. We then write

$$Z_\alpha = \frac{Q_\alpha}{\Lambda^{3N}\, N!} \quad , \quad Q_\alpha = \int_\alpha \mathrm{d}\mathbf{q}\, e^{-\beta V} \quad , \tag{2.6}$$

where the integral $\int_\alpha$ is restricted to the region of configuration space corresponding to the phase α, and $F_\alpha = -k_B T \ln Z_\alpha$ is the free energy of this phase. We assume the density of the crystal is sufficiently high to guarantee the structural stability of the phase, which we define empirically: over the course of any reasonably long simulation, no portion of the crystal will melt or change phase, nor will two or more particles exchange locations within the lattice.

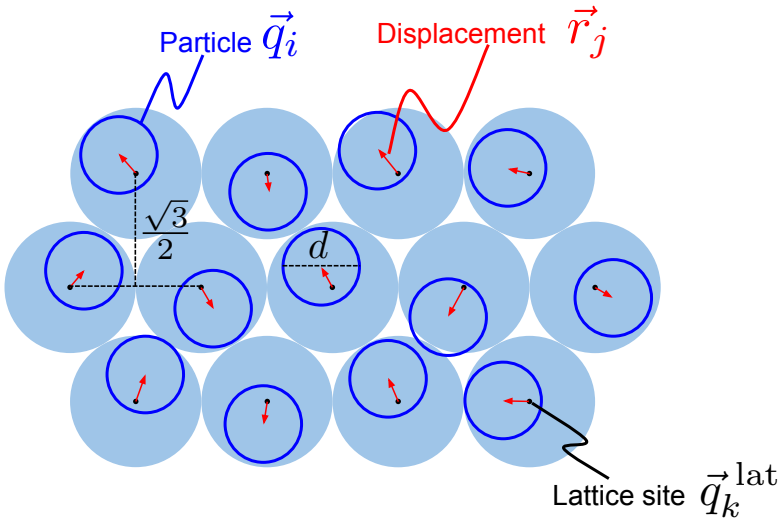


Figure 2.1: The shaded circles of diameter 1 depict a hexagonal layer of a close-packed crystal, with particles arranged at lattice sites $\vec{q}_k^{\,\mathrm{lat}}$. The smaller off-site circles ($d < 1$) represent a conformation at a lower density, with each particle displaced from its lattice site by $\vec{r}_k = \vec{q}_k - \vec{q}_k^{\,\mathrm{lat}}$.

Now let us focus specifically on a crystal composed of hard spheres of diameter d interacting under the pairwise potential (also see Eqs. 1.15 and 1.16),

$$V_{\rm HS}\left(\vec{q}_i,\ \vec{q}_j\right) = \begin{cases} \infty & ,\ |\vec{q}_i - \vec{q}_j| \leq d \\ 0 & ,\ |\vec{q}_i - \vec{q}_j| > d \end{cases} \quad , \tag{2.7}$$

so that the total potential energy is

$$V(\mathbf{q}) = \sum_{i<j} V_{\rm HS}\left(\vec{q}_i,\ \vec{q}_j\right) \quad . \tag{2.8}$$

This potential divides $\mathbf{q}$-space into energetically allowed ($V = 0$) and forbidden ($V = \infty$) regions; the former are further subdivided into contributions from various phases. We wish to compute $Z_{\rm FCC}$ (see Eq. 2.6), subject to these considerations:

(1) The number of particles is $N = n_x \times n_y \times n_z$, indicating n_z layers of $n_x \times n_y$ hexagonally packed spheres, stacked in the z direction. Fig. 2.1 depicts a portion of one layer, in the $x-y$ plane. The numbers n_x, n_y and n_z are chosen to accommodate periodic boundary conditions.

(2) The dimensions of the box $\mathbf{\Omega}$ are taken to be $l_x = n_x$, $l_y = n_y \cdot \sqrt{3}/2$, and $l_z = n_z \cdot \sqrt{6}/3$, hence its volume is $|\mathbf{\Omega}| = N/\sqrt{2}$. These values correspond to a close-packed solid of spheres with unit diameter (see Fig. 2.1). We are interested in the case $d \lesssim 1$, in which the particles have some freedom to move but the FCC structure is stable. We take $\mathbf{\Omega}$ to be centered at the origin, $\vec{0} = (0, 0, 0)$.

(3) The quantity $\rho = N\pi d^3/6|\mathbf{\Omega}|$ specifies the fractional volume of space occupied by the spheres, and has the maximum value $\rho_0 = \pi/\sqrt{18}$ when $d = 1$, that is

when the particles form a close-packed crystal. We define the reduced density [7],

$$\rho_r = \rho/\rho_0 = d^3 \leq 1 \quad , \tag{2.9}$$

and assume $\rho_r > 0.7360$ to prevent melting [32].

2.2.2 Change of coordinates

Among the $3N$ degrees of freedom of our crystal, let us distinguish between three "external" degrees associated with the translation of the entire crystal, and $3N-3$ "internal" degrees that specify an arrangement of particles relative to one another. To this end, we introduce a new set of coordinates to specify the particles' locations. In place of $\mathbf{q} = (\vec{q}_1, \cdots \vec{q}_N) \in \mathbf{\Omega}^N$, we will use

$$\mathbf{x} = (\vec{c}, \vec{r}_2, \cdots \vec{r}_N) \in \omega \times \mathbf{\Omega}^{N-1} \tag{2.10}$$

defined as follows. First, let $\{\vec{\Delta}_1^{\mathrm{FCC}}, \cdots, \vec{\Delta}_N^{\mathrm{FCC}}\}$ denote the locations of the N sites of a reference FCC lattice that is contained within $\mathbf{\Omega}$ and is centered at the origin: $\sum_{i=1}^N \vec{\Delta}_i^{\mathrm{FCC}} = \vec{0}$. The periodic images of the reference lattice form an infinite FCC lattice in $\mathbb{R}^3$. Next, let ω be a small rectangular region within $\mathbf{\Omega}$, of dimensions $(l_x/N) \times (l_y/N) \times (l_z/N)$ and volume $|\omega| = |\mathbf{\Omega}|/N^3$, centered at the origin. For any $\vec{c} \in \omega$, the vectors

$$\vec{q}_i^{\,\mathrm{lat}} = [\vec{\Delta}_i^{\mathrm{FCC}} + \vec{c}]_{\mathbf{\Omega}} \quad , \quad 1 \leq i \leq N \tag{2.11}$$

describe an FCC lattice obtained by displacing the entire reference lattice by $\vec{c}$. Finally, each vector $\vec{r}_i$ specifies the displacement of the i-th particle from its lattice site $\vec{q}_i^{\,\mathrm{lat}}$:

$$\vec{q}_i = \left[\vec{q}_i^{\,\mathrm{lat}} + \vec{r}_i\right]_{\mathbf{\Omega}} = \left[\vec{\Delta}_i^{\mathrm{FCC}} + \vec{c} + \vec{r}_i\right]_{\mathbf{\Omega}} \tag{2.12}$$

(see Fig. 2.1). The vectors $(\vec{r}_2, \cdots \vec{r}_N) \in \mathbf{\Omega}^{N-1}$ are independent variables, and $\vec{r}_1$ is defined by the restriction

$$\sum_{i=1}^{N} \vec{r}_i = \vec{0}\,. \tag{2.13}$$

By varying $\vec{c}$ we translate the entire crystal, and by varying the $\vec{r}_i$'s we rearrange particles relative to one another. Thus $\vec{c}$ and $\{\vec{r}_i\}$ are the external and internal degrees of freedom, respectively. Note that the potential V depends on $\{\vec{r}_i\}$ but not on $\vec{c}$.

Here we establish that these coordinate sets ($\mathbf{q}$ and $\mathbf{x}$) are in one-to-one correspondence – in other words, any arrangement of N particles in the box $\mathbf{\Omega}$ can be represented equally well by $\mathbf{x}$ as by $\mathbf{q}$ – and we calculate the Jacobian $|\partial\mathbf{q}/\partial\mathbf{x}|$.

The transformation from $\mathbf{x}$ to $\mathbf{q}$ was provided by Eq. 2.12. We denote this transformation as a mapping $\mathcal{M} : \mathbf{x} \to \mathbf{q}$, whose inverse $\mathcal{M}^{-1} : \mathbf{q} \to \mathbf{x}$ is given by

$$\vec{c} \;=\; \frac{1}{N} \cdot \left[\sum_{i=1}^{N} \vec{q}_i\right]_{\mathbf{\Omega}} \tag{2.14a}$$

$$\vec{r}_i \;=\; \left[\vec{q}_i - \vec{\Delta}_i - \vec{c}\right]_{\mathbf{\Omega}} \quad , \quad 2 \le i \le N \quad . \tag{2.14b}$$

Eq. 2.14 allows us to determine the external and internal degrees of freedom, $\mathbf{x}$, given the locations of all the particles, $\mathbf{q}$. One can verify by inspection that $\mathcal{M}^{-1}(\mathcal{M}(\mathbf{x})) =$

$\mathbf{x}$ for any $\mathbf{x} \in \omega \times \mathbf{\Omega}^{N-1}$, and $\mathcal{M}\left(\mathcal{M}^{-1}(\mathbf{q})\right) = \mathbf{q}$ for any $\mathbf{q} \in \mathbf{\Omega}^N$.

To calculate the Jacobian of the mapping $\mathcal{M}$, we first write the matrix $\partial\mathbf{q}/\partial\mathbf{x}$ in block-diagonal form, by separately grouping the contributions from the x, y, and z components:

$$\frac{\partial \mathbf{q}}{\partial \mathbf{x}} = \begin{pmatrix} \mathbf{J}_x & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{J}_y & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{J}_z \end{pmatrix} \quad . \tag{2.15}$$

From Eq. 2.12 we have

$$\mathbf{J}_x = \mathbf{J}_y = \mathbf{J}_z = \begin{pmatrix} 1 & -1 & -1 & -1 & \cdots & -1 \\ 1 & 1 & 0 & 0 & \cdots & 0 \\ 1 & 0 & 1 & 0 & \cdots & 0 \\ 1 & 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & 0 & \cdots & 1 \end{pmatrix}_{N \times N} \quad . \tag{2.16}$$

The determinant of this matrix is N, as computed using the expansion of minors. Since $\partial\mathbf{q}/\partial\mathbf{x}$ is block-diagonal, we obtain

$$\left| \frac{\partial \mathbf{q}}{\partial \mathbf{x}} \right| = |\mathbf{J}_x| \cdot |\mathbf{J}_y| \cdot |\mathbf{J}_z| = N^3 \quad . \tag{2.17}$$

Consistency check. Since the Jacobian $|\partial\mathbf{q}/\partial\mathbf{x}|$ is independent of $\mathbf{x}$, it should be equal to the ratio between the volumes of the domain and the range of $\mathcal{M}$. The domain of $\mathcal{M}$ is $\mathbf{x}$-space, whose volume is $|\omega| \times |\mathbf{\Omega}|^{N-1} = |\mathbf{\Omega}|^N / N^3$; the range of $\mathcal{M}$

is **q**-space, whose volume is $|\mathbf{\Omega}|^N$. Dividing the latter by the former we get N^3, in agreement with Eq. 2.17.

2.2.3 Evaluation of the partition function

As shown above, the coordinate sets **q** and **x** are in one-to-one correspondence, and the Jacobian for the transformation between them is $|\partial\mathbf{q}/\partial\mathbf{x}| = N^3$. This allows us to integrate out the external degrees (associated with translational symmetry) in Eq. 2.5:

$$Q = \int_{\mathbf{\Omega}^N} \mathrm{d}\mathbf{q}\, e^{-\beta V} = |\mathbf{\Omega}| \int_{\mathbf{\Omega}} \mathrm{d}\vec{r}_2 \cdots \int_{\mathbf{\Omega}} \mathrm{d}\vec{r}_N\, e^{-\beta V} \tag{2.18}$$

where we used $|\omega| = |\mathbf{\Omega}|/N^3$. Introducing the shorthand notation

$$\int' \mathrm{d}\mathbf{r} \equiv \int_{\mathbb{R}^3} \mathrm{d}\vec{r}_1 \int_{\mathbf{\Omega}} \mathrm{d}\vec{r}_2 \cdots \int_{\mathbf{\Omega}} \mathrm{d}\vec{r}_N\, \delta^3\left(\sum_{i=1}^{N} \vec{r}_i\right) \tag{2.19}$$

where $\mathbf{r} = (\vec{r}_1, \cdots \vec{r}_N)$, Eq. 2.18 becomes

$$Q = |\mathbf{\Omega}| \int' \mathrm{d}\mathbf{r}\, e^{-\beta V} \tag{2.20}$$

This result gives Q as an integral over the N displacement vectors $\{\vec{r}_i\}$, and the δ^3-function enforces Eq. 2.13.

We will use the term *conformation* to denote a set of displacement vectors $\{\vec{r}_i\}$ that satisfy Eq. 2.13, with $\vec{r}_i \in \mathbf{\Omega}$ for $2 \leq i \leq N$. Thus, the notation $\int' \mathrm{d}\mathbf{r}$ indicates an integral over $(3N-3)$-dimensional conformation space, equivalently over the internal degrees of freedom. To compute Q_{FCC} we restrict the integral in

Eq. 2.20 to FCC conformations:

$$Q_{\mathrm{FCC}} = |\mathbf{\Omega}| \int_{\mathrm{FCC}}' \mathrm{d}\mathbf{r}\, e^{-\beta V} \quad . \tag{2.21}$$

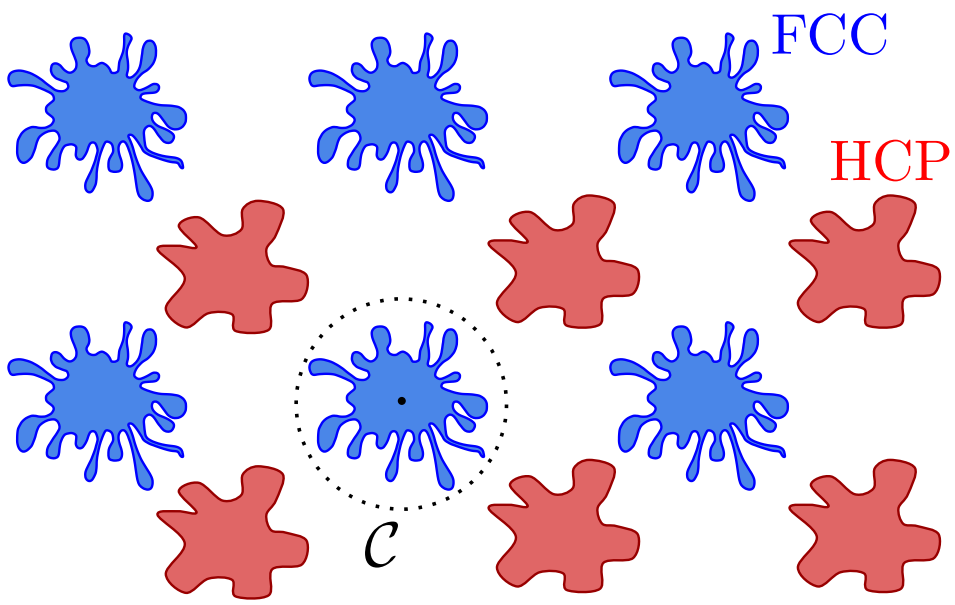


Figure 2.2: In this sketch of conformation space, the $(3N-3)$ - dimensional islands represent structurally stable phases, such as FCC (ink stains) and HCP (jigsaw pieces), with $N!$-fold degeneracy due to permutations of N indistinguishable particles. The central FCC island, $\mathcal{C}$, contains the origin $\mathbf{0} = (0, \cdots 0)$.

Let us now visualize the energy landscape in conformation space. For a *close-packed* hard sphere crystal (i.e. $d = 1$), there are a finite number of isolated, energetically allowed conformations. These correspond to exact FCC and HCP lattices, as well as other close-stacked arrangements, or *polytypes* [33,34]. If we now decrease the particle diameters slightly ($d \lesssim 1$), to give the particles room to move about, then each of these isolated points in conformation space grows into an energetically allowed, structurally stable *island*, as sketched in Fig. 2.2.

Our assumption of structural stability guarantees that these islands are well

separated. In a simulation performed at fixed $\vec{c}$ to prevent drift (see Sec. 2.4), the system remains confined within a single island, as each particle remains close to a single, fixed lattice site.

There are a total of $N!$ stable FCC islands, which are identical apart from permutations of the particles among the lattice sites. We will use the term *central island*, $\mathcal{C}$, to specify the FCC island that includes the origin $\mathbf{r} = (\vec{0}, \cdots, \vec{0}) \equiv \mathbf{0}$. For conformations in this island, each particle is located near its lattice site: $\vec{q}_i \approx \vec{q}_i^{\,\mathrm{lat}}$ (see Eq. 2.12).

Combining Eqs. 2.6 and 2.21 with the $N!$-fold degeneracy of FCC islands gives us

$$Z_{\mathrm{FCC}} = \frac{|\mathbf{\Omega}|}{\Lambda^{3N}} \int_{\mathcal{C}}' \mathrm{d}\mathbf{r}\, e^{-\beta V} \tag{2.22}$$

(For a hard sphere crystal the factor $e^{-\beta V}$ is redundant, since $V = 0$ for any $\mathbf{r} \in \mathcal{C}$. However, Eq. 2.22 and the results that follow remain valid for more general particle interactions, therefore we will retain this factor.) The integral in Eq. 2.22 is the volume of the central island, $\mathcal{C}$, in conformation space. We now describe how to compute this volume using simulations that sample the central island.

We will use $r = |\mathbf{r}|$ as a scalar order parameter with a natural interpretation: it is a measure of the net "distance" of the N particles from the reference lattice. Rewriting Eq. 2.22 in terms of an integral along this order parameter, we have

$$Z_{\mathrm{FCC}} = \frac{|\mathbf{\Omega}|}{\Lambda^{3N}} \int_0^\infty \mathrm{d}r\, \pi(r) \tag{2.23}$$

where

$$\pi(r) \equiv P_0(r)s(r) \tag{2.24}$$

$$P_0(r) = \frac{\int' \mathrm{d}\mathbf{r}\, \Theta_{\mathcal{C}}\, e^{-\beta V}\, \delta(r - |\mathbf{r}|)}{\int' \mathrm{d}\mathbf{r}\, \delta(r - |\mathbf{r}|)} \equiv \left\langle \Theta_{\mathcal{C}}\, e^{-\beta V} \right\rangle_r \tag{2.25}$$

$$s(r) = \int' \mathrm{d}\mathbf{r}\, \delta(r - |\mathbf{r}|) = \frac{\kappa_n r^{n-1}}{N^{3/2}} \quad . \tag{2.26}$$

(See Appendix. A) For simplicity, we have introduced $n = 3N - 3$ and

$$\kappa_n = \frac{n\pi^{n/2}}{\Gamma(n/2+1)} \quad . \tag{2.27}$$

We have also defined $\Theta_{\mathcal{C}}(\mathbf{r})$ as an indicator whose value is 1 if $\mathbf{r} \in \mathcal{C}$, and 0 if $\mathbf{r} \notin \mathcal{C}$.

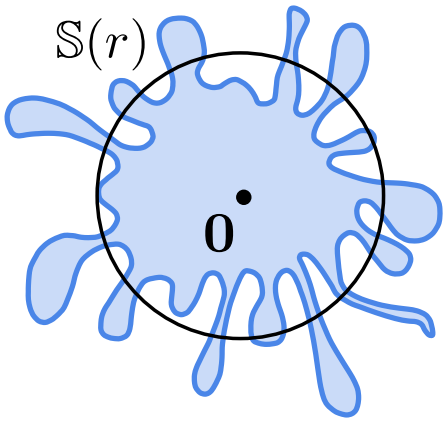


Figure 2.3: The central FCC island $\mathcal{C}$. At the origin, each particle is located at its lattice site, $\vec{q}_k^{\,\mathrm{lat}}$, but with increasing distance from the origin, the fraction of energetically allowed conformations decreases rapidly. The circle $\mathbb{S}(r)$ represents a hypersphere of radius r.

For a hard-sphere crystal, Eqs. 2.23-2.26 have simple interpretations. Let $\mathbb{S}(r)$ be the set of all conformations with $|\mathbf{r}| = r$. This defines a hyperspherical surface of radius r in n-dimensional conformation space: $s(r)$ is (effectively) its surface

area, and $P_0(r)$ is the fraction of this surface that belongs to the central island $\mathcal{C}$. The integral of their product, in Eq. 2.23, gives the volume of $\mathcal{C}$. Note that $P_0(0) = 1$, since $V(\mathbf{q}) = 0$ when each $\vec{q}_i = \vec{q}_i^{\,\text{lat}}$. The value of $P_0(r)$ decays rapidly with increasing r, due to particle-particle overlaps. Fig. 2.3 schematically depicts $\mathbb{S}(r)$ as a circle of radius r: as r increases the fraction of the circle's perimeter within $\mathcal{C}$ decreases. Thus $\pi(r)$ in Eq. 2.23 is the product of a decreasing function, P_0, and an increasing one, s, resulting in a peak around a value r_0, as sketched in Fig. 2.4. To evaluate Z_{FCC}, we must compute the area underneath this peak.

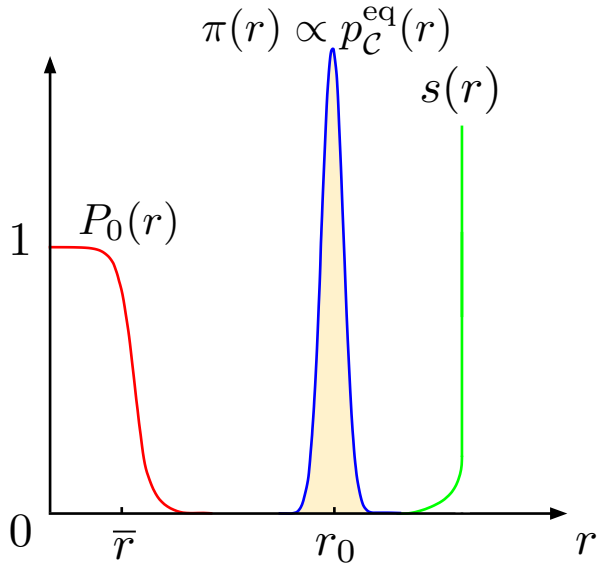


Figure 2.4: $P_0(r)$, $s(r)$ and $\pi(r)$, sketched on the same scale for convenience. $\pi(r)$ is proportional to the equilibrium distribution of the order parameter, $p_{\mathcal{C}}^{\text{eq}}(r)$. The area underneath $\pi(r)$ is the integral in Eq. 2.23.

For a more general $V(\mathbf{q})$ (not hard sphere), Eqs. 2.23 - 2.26 remain valid but the interpretation of $P_0(r)$ is modified, as the Boltzmann factor is no longer necessarily 0 or 1.

Now note that the normalized equilibrium probability distribution of the order

parameter sampled within $\mathcal{C}$ is given by

$$p_{\mathcal{C}}^{\mathrm{eq}}(r) = \frac{\int_{\mathcal{C}}' \mathrm{d}\mathbf{r}\, e^{-\beta V}\, \delta(r - |\mathbf{r}|)}{\int_{\mathcal{C}}' \mathrm{d}\mathbf{r}\, e^{-\beta V}} = \frac{\pi(r)}{\int_{\mathcal{C}}' \mathrm{d}\mathbf{r}\, e^{-\beta V}} \quad . \tag{2.28}$$

Comparing Eqs. 2.22 and 2.28, we get

$$Z_{\mathrm{FCC}} = \frac{|\mathbf{\Omega}|}{\Lambda^{3N}} \frac{\pi(\bar{r})}{p_{\mathcal{C}}^{\mathrm{eq}}(\bar{r})} = \frac{|\mathbf{\Omega}|}{\Lambda^{3N}} \frac{\kappa_n \bar{r}^{n-1}}{N^{3/2}} \frac{P_0(\bar{r})}{p_{\mathcal{C}}^{\mathrm{eq}}(\bar{r})} \tag{2.29}$$

for any $\bar{r}$ such that $p_{\mathcal{C}}^{\mathrm{eq}}(\bar{r}) \neq 0$. Alternatively, Eq. 2.29 can be derived from Eq. 2.23 using the relation

$$\int_0^\infty \mathrm{d}r\, \pi(r) = \frac{\pi(\bar{r})}{p_{\mathcal{C}}^{\mathrm{eq}}(\bar{r})} \int_0^\infty \mathrm{d}r\, p_{\mathcal{C}}^{\mathrm{eq}}(r) = \frac{\pi(\bar{r})}{p_{\mathcal{C}}^{\mathrm{eq}}(\bar{r})} \tag{2.30}$$

which follows from the proportionality between $\pi(r)$ and $p_{\mathcal{C}}^{\mathrm{eq}}(r)$ (see Eq. 2.28). With Eq. 2.29, the computation of Z_{FCC} reduces to a value matching operation: we must compute both P_0 and $p_{\mathcal{C}}^{\mathrm{eq}}$ at some value $\bar{r}$. Note, however, that there is little overlap between these two functions, as illustrated in Fig. 2.4.

We compute $P_0(\bar{r})$ directly, by sampling points $\mathbf{r}$ uniformly from the hypersphere $\mathbb{S}(\bar{r})$ (see Sec. 2.4), and evaluating the fraction of the sampled conformations for which $V = 0$. For this to succeed, we must choose $\bar{r}$ small enough that $P_0(\bar{r})$ is non-negligible. For large N, this value $\bar{r}$ is located deep in the left tail of $p_{\mathcal{C}}^{\mathrm{eq}}$ (Fig. 2.4). In order to evaluate $p_{\mathcal{C}}^{\mathrm{eq}}(\bar{r})$ at this small value of $\bar{r}$, we then use biased (umbrella) sampling, combined with the Weighted Histogram Analysis Method (WHAM) [35].

Specifically, we perform a set of biased Monte Carlo simulations, at fixed $\vec{c}$, using potential energy functions

$$V_\lambda(\mathbf{r}) = V(\mathbf{r}) + \lambda \ln(r) \quad , \tag{2.31}$$

at discrete values of λ ranging from 0 to $n-1$. Here, $V(\mathbf{r})$ is the unbiased potential (Eq. 2.8) expressed in terms of $\mathbf{r}$. For each λ we collect a histogram of the order parameter r, as illustrated in Fig. 2.5. At $\lambda = 0$ this histogram is just the unbiased distribution $p_{\mathcal{C}}^{\text{eq}}(r)$, which is peaked at $r = r_0$. With increasing λ, the histogram shifts toward lower values of r. By applying the WHAM algorithm, $p_{\mathcal{C}}^{\text{eq}}(r)$ can be reconstructed with good statistics all the way down to $r = \bar{r}$. From the values of $P_0(\bar{r})$ and $p_{\mathcal{C}}^{\text{eq}}(\bar{r})$ we finally obtain Z_{FCC} (Eq. 2.29).

2.3 Result

To test our method, we applied it to FCC and HCP hard sphere crystals of various sizes ($N = 96$ to 1728), at reduced density $\rho_r = 0.7778$. We computed the free energy per particle, $f_\alpha = -k_B T N^{-1} \ln Z_\alpha$, setting $\Lambda = 1$ for convenience. Table 2.1 gives our results for f_{FCC}, f_{HCP} and $\Delta f \equiv f_{\text{HCP}} - f_{\text{FCC}}$, with the system size specified by the stacking (n_z, n_y, n_z). This table also shows the results of Ref. [1] (in boldface), where Δf was computed directly with the lattice switch method, without separately computing f_{FCC} and f_{HCP}. We find good agreement (see also Fig. 2.7), which provides evidence of the accuracy of our method, particularly since f_{FCC} and f_{HCP} are nearly equal.

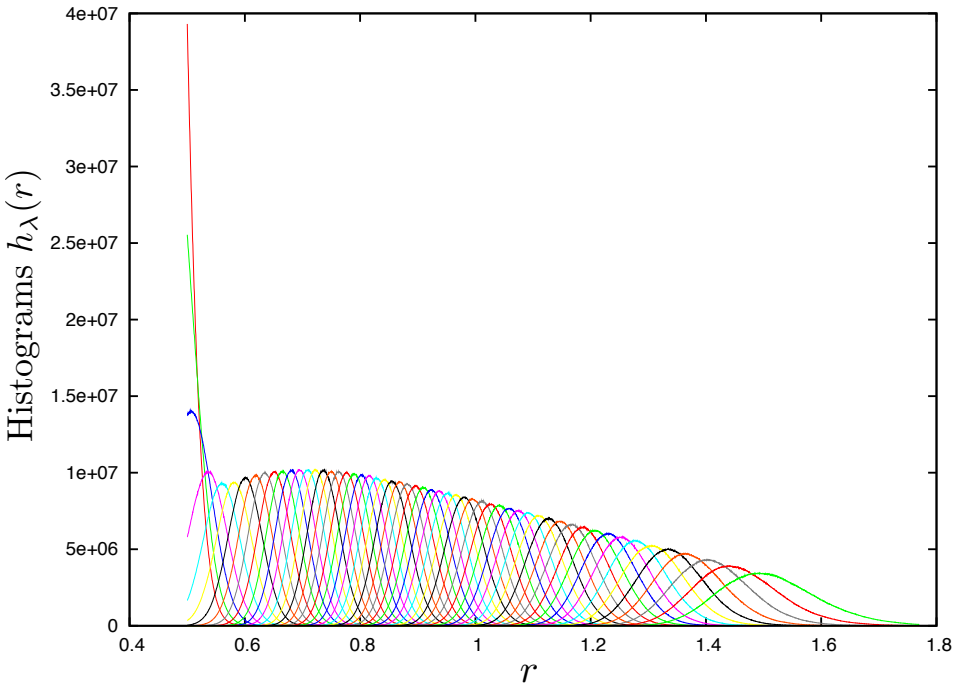


Figure 2.5: Fifty-six histograms of the order parameter r, generated from biased sampling at different values of λ, for a $6 \times 6 \times 6$ FCC crystal with reduced density $\rho_r = 0.7778$. For computational efficiency, we apply a hard constraint so as to sample only the region $r > \bar{r} = 0.5$ (see Sec. 2.4.2).

We separated our results into two groups according to the number of stacked hexagonally packed layers, $n_z = 6$ and $n_z = 12$, varying (n_x, n_y) within each group. In Fig. 2.6 both f_{FCC} and f_{HCP} are plotted against the total particle number, N, and we see that the FCC free energy is consistently slightly lower than the HCP free energy. In Fig. 2.7 we plot the difference Δf, which appears to saturate at $\Delta f \approx 0.0013\ k_B T$ for 6 layers and $\Delta f \approx 0.0009\ k_B T$ for 12 layers. This suggests a finite size effect that is not surprising given the relatively small number of stacked layers.

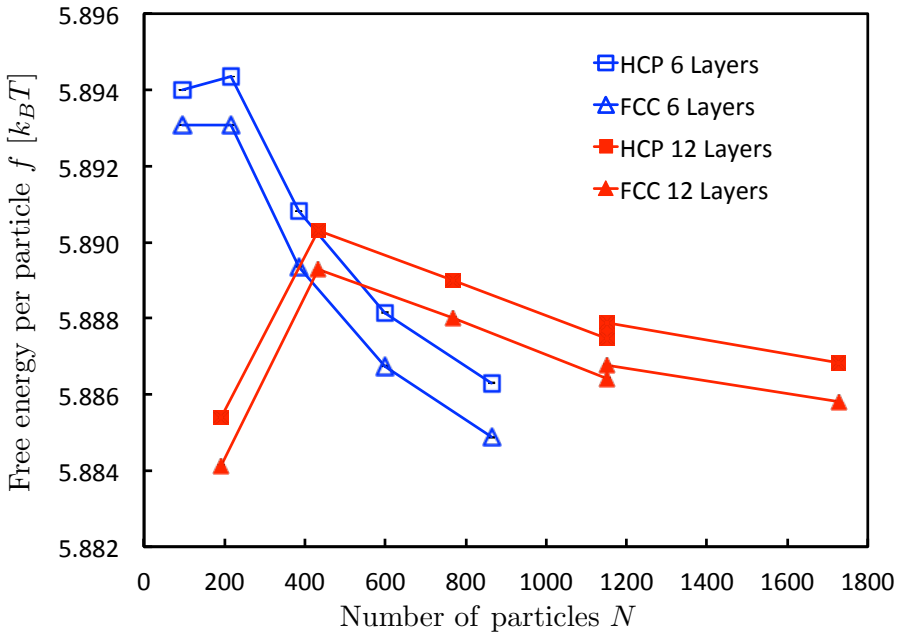


Figure 2.6: Free energy per particle for HCP (squares) and FCC (triangles) hard sphere crystals, with number of layers $n_z = 6$ and $n_z = 12$. Error bars, not shown, are much smaller than symbol size.

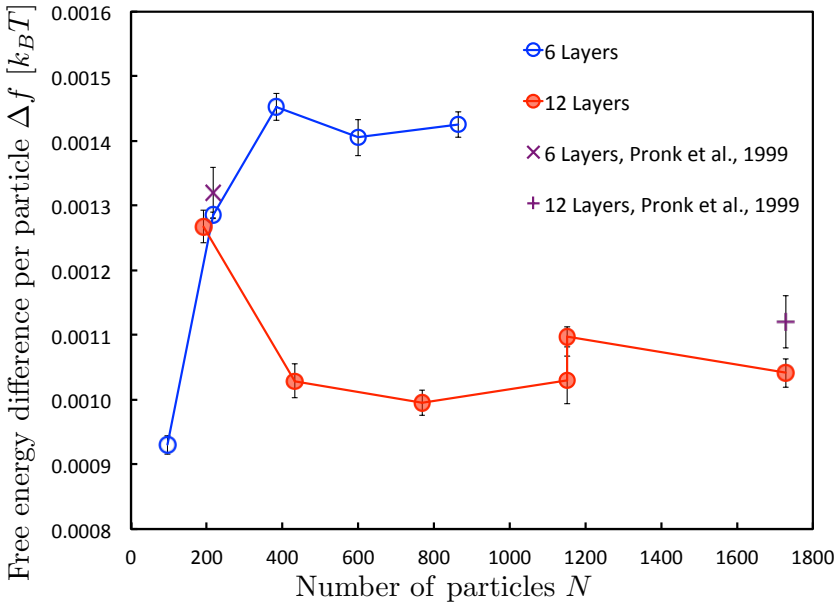


Figure 2.7: Free energy difference per particle, $\Delta f = f_{\mathrm{HCP}} - f_{\mathrm{FCC}}$, calculated using our method for $n_z = 6$ and $n_z = 12$. Error bars, corresponding to one standard deviation, are estimated using the WHAM method. The crosses indicate results obtained by the lattice switch method, taken from Ref. [1].

Table 2.1: Summary of results. Calculated free energies and free energy differences, per particle, in units of k_BT. For the results of the present work, standard deviations obtained using WHAM are given in parentheses in units of $10^{-6}k_BT$. Results of Ref. [1] are shown in **boldface**, where the units of the uncertainties (in parentheses) are $10^{-5}k_BT$.

(n_x, n_y, n_z)	f_{HCP}	f_{FCC}	Δf
(4,4,6)	5.894002(10)	5.893072(10)	0.000930(14)
(6,6,6)	5.894353(03)	5.893067(04)	0.001286(05)
(6,6,6)	–	–	**0.00132(4)**
(8,8,6)	5.890813(14)	5.889361(15)	0.001453(21)
(10,10,6)	5.888151(19)	5.886746(21)	0.001405(28)
(12,12,6)	5.886297(13)	5.884872(15)	0.001425(20)
(4,4,12)	5.885403(18)	5.884135(18)	0.001268(26)
(6,6,12)	5.890299(18)	5.889270(19)	0.001029(26)
(8,8,12)	5.889013(13)	5.888018(14)	0.000995(19)
(12,8,12)	5.887460(25)	5.886430(27)	0.001030(37)
(8,12,12)	5.887874(11)	5.886776(11)	0.001097(16)
(12,12,12)	5.886842(15)	5.885801(17)	0.001041(22)
(12,12,12)	–	–	**0.00112(4)**

2.4 Numerical methods

2.4.1 Direct estimates of $P_0(\bar{r})$ and $p_{\mathcal{C}}^{\mathrm{eq}}(\bar{r})$.

To estimate $P_0(\bar{r})$, we generate uniform samples from a hyperspherical surface of radius $\bar{r}$ in conformation space, as follows (also see Ref. [36]). We first generate $\mathbf{n} = (\vec{n}_1, \cdots, \vec{n}_N)$ by sampling from $3N$ independent normal (Gaussian) distributions with unit variance and zero mean. We then define $\mathbf{n}^0 = (\vec{n}_1^0, \cdots, \vec{n}_N^0)$ where $\vec{n}_i{}^0 = \vec{n}_i - \sum_{j=1}^{N} \vec{n}_j/N$. Finally, we set $\mathbf{r} = \bar{r} \cdot \mathbf{n}^0/|\mathbf{n}^0|$. This point satisfies both $|\mathbf{r}| = \bar{r}$ and Eq. 2.13, and is a random, uniform sample from the hypersphere $\mathbb{S}(r)$. We use Eq. 2.12 to translate $\mathbf{r}$ into $\mathbf{q}$ (setting $\vec{c} = \vec{0}$ for convenience) and we check for overlaps between pairs of spheres. $P_0(\bar{r})$ is the fraction of samples for which

there are no overlaps.

To determine $p_{\mathcal{C}}^{\mathrm{eq}}(\bar{r})$, we perform biased Metropolis Monte Carlo simulations, using the potentials in Eq. 2.31. These simulations produce a set of histograms $\{h_\lambda(r)\}$, from $\lambda = 0$ to $n-1$, with the values of λ chosen to produce good overlaps between adjacent histograms; see Fig. 2.5 for an illustration. Note that $h_0(r) \propto p_{\mathcal{C}}^{\mathrm{eq}}(r)$; while $h_{n-1}(r) \propto \pi(r)/s(r) = P_0(r)$, as the biasing potential $(n-1)\ln r$ cancels the geometrical factor $s(r) \propto r^{n-1}$. We use the Weighted Histogram Analysis Method (WHAM) [35] to combine all the histograms, producing a distribution $p_{\mathcal{C}}^{\mathrm{eq}}(r)$ that is accurate down to $r = \bar{r}$.

The simulations described in the previous paragraph are performed at fixed $\vec{c}$, to prevent drift. Specifically, each simulation is initialized in the conformation $\mathbf{r} = \mathbf{0}$, with each particle located at its lattice site: $\vec{q}_i = \vec{q}_i^{\,\mathrm{lat}}$. A trial move is generated by applying equal and opposite displacements, $\pm\vec{\delta}$, to $\vec{r}_i$ and $\vec{r}_j$, a randomly chosen pair of particles, with $\vec{\delta}$ sampled symmetrically to satisfy detailed balance. The move is then accepted or rejected according the Metropolis rule. By sampling in this manner and after allowing for relaxation to equilibrium, we generate a set of conformations from which we construct the histogram $h_\lambda(r)$.

In the following, we discuss the choice of $\bar{r}$ as well as an efficient (and approximated) method for estimating $P_0(\bar{r})$ for large N.

2.4.2 Efficient estimation of $P_0(\bar{r})$.

We have described numerical procedures to estimate $P_0(\bar{r})$ and $p_{\mathcal{C}}^{\mathrm{eq}}(\bar{r})$. A judicious choice of $\bar{r}$ can reduce the computational effort of obtaining these quantities, and for large N the value of $P_0(\bar{r})$ can be estimated efficiently by extrapolating from smaller simulations.

The choice of $\bar{r}$ involves a tradeoff. On the one hand, when computing $P_0(\bar{r})$ by direct sampling of the surface $\mathbb{S}(\bar{r})$, it is favorable to choose a small value $\bar{r}$ so that $P_0(\bar{r})$ is not negligible (see Fig. 2.4). On the other hand, when using biased Monte Carlo sampling to compute $p_{\mathcal{C}}^{\mathrm{eq}}(\bar{r})$, larger values of $\bar{r}$ are preferable, since the trial move acceptance probability becomes very small in the region $r \approx 0$, due to the strong biasing potential $(n-1)\ln r$. We can partially avoid this low acceptance region by sampling only in the range $r > \bar{r}$, in other words by rejecting all trial moves that reduce the value of r below $\bar{r}$.

As a particular example, for the case $N = 216$ we chose $\bar{r} = 0.5$. For this choice we have $P_0(0.5) \approx 0.1$, which is readily estimated using direct sampling from the surface $\mathbb{S}(0.5)$. The estimation of $p_{\mathcal{C}}^{\mathrm{eq}}(r)$ is illustrated in the main test (see Fig. 2.5), where overlapping biased histograms $h_\lambda(r)$ are obtained all the way down to $r = 0.5$.

Having calculated $P_0(0.5) = 0.0950$ using direct sampling (Sec. 2.4.1) for $N = 216$, we used a scaling relation (Eq. 2.32) to extrapolate to larger numbers of particles. In what follows, we denote the function $P_0(r)$ corresponding to system size N as $P_0(r; N)$, to make explicit its dependence on the number of particles.

Consider a conformation $\mathbf{r} = (\vec{r}_1, \cdots, \vec{r}_N)$ sampled randomly and uniformly from the hyperspherical surface $\mathbb{S}(\bar{r})$. Since $\sum_{i=1}^{N} |\vec{r}_i|^2 = \bar{r}^2$, the typical distance of each particle from its lattice site is $\bar{a} \equiv \bar{r}/\sqrt{N}$. When a is much smaller than the lattice spacing, the probability q of overlap between a given pair of neighboring particles is very small: $q \ll 1$. The probability of *no* overlap between *any* pairs of particles can then be estimated as $P_0 \approx (1-q)^{Nb/2}$, where b is the number of nearest neighbors per particle. For sufficiently large N, if we increase both N and $\bar{r}$ while holding $\bar{a} = \bar{r}/\sqrt{N}$ fixed, then q remains constant and

$$P_0(\sqrt{N}\bar{a}; N) \approx (1-q)^{Nb/2} \equiv \epsilon^N. \tag{2.32}$$

From this scaling relation we obtain

$$P_0\left(0.5\sqrt{\frac{N}{216}}\,;\, N\right) \;\approx\; P_0\,(0.5\,;\,216)^{N/216} = 0.0950^{N/216}. \tag{2.33}$$

As a consistency check, we computed $P_0(0.5\sqrt{N/216}; N)$ by direct sampling for $N = 96$, 192, and 288, and found excellent agreement with Eq. 2.33 (shown in Fig. 2.8).

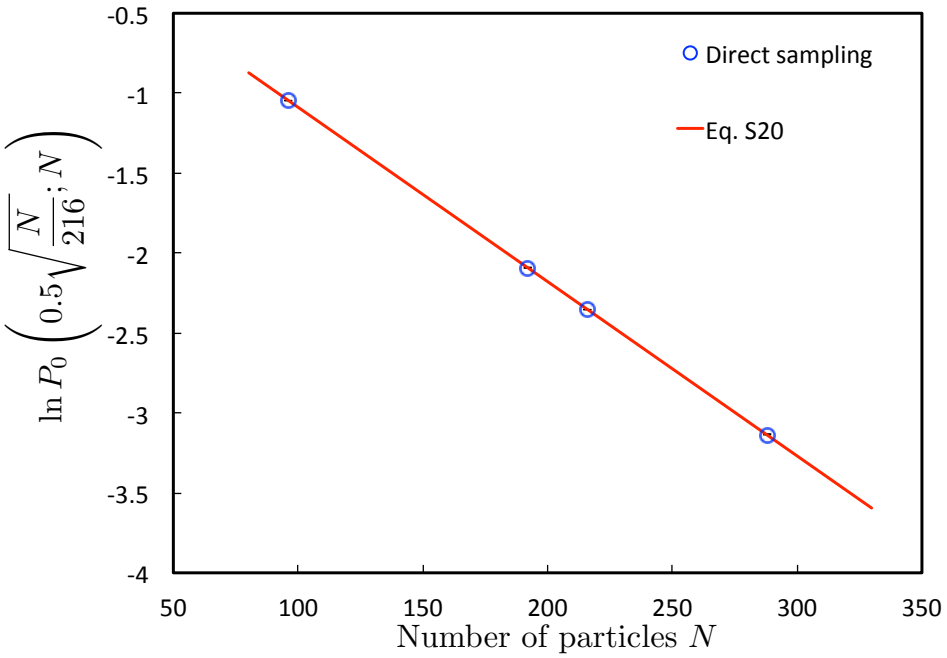


Figure 2.8: The four circles show the values of $\ln P_0(0.5\sqrt{N/216}; N)$ obtained by directly sampling FCC hard sphere crystals at reduced density $\rho_r = 0.7778$, for $N = 96$, 192, 216 and 288. The statistical errors, represented by the tiny error bars within the circles, are smaller than the symbol size. The straight line shows Eq. 2.33.

Chapter 3: Free Energies of Crystalline Solids in the Thermodynamic Limit

In Sec. 2.4.2 we have claimed and numerically verified a scaling property (Eq. 2.32) that is valid for small values of r. In this chapter, we focus on various scaling properties in the thermodynamics limit, i.e. when the number of particles increases ($N \gg 1$) at fixed density ρ. First, we show that in this limit, the scaling relation (Eq. 2.32) for hard sphere crystals is valid not only for small values of $\bar{r}$. Furthermore, we derive an asymptotic formula for the crystalline free energy per particle in the thermodynamic limit, which is valid for more general crystals (e.g. soft interaction potentials).

3.1 Scaling property of $P_0(r; N)$ for hard sphere crystals

Recall from Sec. 2.2.3 that $P_0(r; N)$ is an average of the Boltzmann factor $e^{-\beta V}$ by uniformly sampling the set of displacement vectors that satisfy $|\mathbf{r}| = r$. Alternatively, we can denote $P_0(r; N)$ as $P_0(\sqrt{N}a; N)$, by introducing

$$a := \sqrt{\frac{|\vec{r}_1|^2 + |\vec{r}_2|^2 + \cdots + |\vec{r}_N|^2}{N}} = \sqrt{\frac{|\mathbf{r}|^2}{N}} = \frac{r}{\sqrt{N}} \tag{3.1}$$

where a is the root mean squared (r.m.s) displacement of all particles from their lattice sites. Intuitively, $P_0(\sqrt{N}a; N)$ is the probability of obtaining an allowed configuration, when sampling randomly from all configurations that satisfy the constraint $\sqrt{|\mathbf{r}|^2/N} = a$.

In this section, our discussion is restricted to hard sphere crystals. Note that, for a hard sphere crystal at finite temperature, the Boltzmann factor $e^{-\beta V_{\rm HS}}$ computed for any given configuration can only take two possible values. When no particle overlaps, the configuration is energetically allowed ($e^{-\beta V_{\rm HS}} = 1$), and when one or more pairs of particles overlap, the configuration is not allowed ($e^{-\beta V_{\rm HS}} = 0$). For proposes of illustration, let us consider the direct sampling of configurations as a series of Bernoulli trials [37]. Bernoulli trials are experiments where there are only two possible outcomes (true or false) from each trial. The probability of true (or false) for each experiment is independent and identically distributed. When sampling configurations of the hard sphere, the probability of obtaining "true" (when the configuration is allowed) is $P_0(\sqrt{N}a; N)$, and the probability of obtaining "false" (when the configuration is not allowed) is $1 - P_0(\sqrt{N}a; N)$. In the following discussion, we call the probability of obtaining "true" from a trial the Bernoulli probability. We emphasize that operationally, there are two steps for each trial: The first step is to randomly generate displacement vectors ($\vec{r}_i$'s) for the N particles so that the r.m.s displacement from lattice sites is equal to a. The second step is to check for particle-particle overlap among all particles.

In the thermodynamic limit ($N \gg 1$ at fixed density ρ), we can separate the total system into two approximately independent halves (subsystems), where each

subsystem contains $N/2$ particles with the same density ρ. After dividing the N particles into two independent subsystems, we can perform two independent Bernoulli trials, one for each subsystem, as follows. We first generate a configuration for the subsystem under the restriction that the r.m.s displacement of its $N/2$ particles is equal to a. Next we check whether the configuration of the subsystem is energetically allowed (true or false). For each subsystem, the Bernoulli probability is equal to by $P_0(\sqrt{N/2}\ a; N/2)$.

Now consider a combined experiment of the two subsystems described as follows. Two independent Bernoulli trials are performed on the two subsystems, and if both Bernoulli trials give "true" (i.e. both configurations of the subsystems are energetically allowed), the outcome of the combined experiment is "true". We denote the probability of obtaining "true" from the combined experiment by the *joint Bernoulli experiment*

$$P(\text{joint}) = P_0(\sqrt{N/2}\ a; N/2) \cdot P_0(\sqrt{N/2}\ a; N/2) \quad . \tag{3.2}$$

We claim that in the thermodynamic limit, the Bernoulli probability of the composite system (with N particles) is approximately equal to the joint Bernoulli probability of those two independent subsystems:

$$P_0(\sqrt{N}a; N) \approx P(\text{joint}) = \Big(P_0(\sqrt{N/2}\ a; N/2)\Big)^2 \tag{3.3}$$

This claim includes two approximations for the two steps of a Bernoulli trial: In

the sampling step, we replace the set of generated configurations for the composite system at r.m.s displacement a by a set of joint configurations of subsystems whose r.m.s displacements are both equal to a. The justification of this first approximation is discussed in detail later. In the second step, we ignore the boundary effect between the subsystems. In other words, in a large enough composite system, the interaction between its two subsystems can approximately be ignored. This approximation is valid in the thermodynamic limit.

Now let us focus on the first approximation, by relating it to an ideal gas. Let us consider an N-particle ideal gas equilibrated at temperature T in 3 dimensional space. In the thermodynamic limit, the distribution of the average energy (average kinetic energy) becomes sharply peaked (a delta function), at the average (kinetic) energy

$$\bar{E} = \frac{|\mathbf{p}|^2}{2mN} = \frac{1}{N}\sum_{i=1}^{N}\frac{|\ \vec{p}_i\ |^2}{2m} = \frac{3k_BT}{2} \tag{3.4}$$

where m is the mass of each ideal gas particle, $\vec{p}_i$ is the momentum of the ith particle, and $\mathbf{p}$ denotes the $3N$-dimensional momenta of the ideal gas. By comparing Eq. 3.4 with Eq. 3.1, we find that the ideal gas can be used as an auxiliary system for generating configurations $\mathbf{r}$ satisfying $\sqrt{|\mathbf{r}|^2/N} = a$ in the Bernoulli trials. More specifically, if we take a snapshot of an ideal gas equilibrated at temperature

$$T(a) = \frac{a^2}{3k_Bm} \tag{3.5}$$

and extract its momenta $\mathbf{p}$, then by setting $\mathbf{r} = \mathbf{p}$, the configuration $\mathbf{r}$ immediately

satisfies $\sqrt{|\mathbf{r}|^2/N} = a$. Using the momenta of the ideal gas as an auxiliary, we can justify the validity of the first approximation. In the thermodynamic limit, if the N particle ideal gas is separated into two subsystems, the energy and the temperature of the two subsystems are expected to be the same. In other words, when $N \gg 1$ the set of momenta of a N-particle ideal gas equilibrated at $T(a)$ cannot be distinguished from the set of momenta obtained by combining the momenta from two $(N/2)$-particle ideal gases that are separately prepared at temperature $T(a)$. This justifies our approximation that, when $N \gg 1$, the configurations $\mathbf{r}$ satisfying $\sqrt{|\mathbf{r}|^2/N} = a$ for the N-particle system can be alternatively generated by jointly sampling configurations of its subsystems at r.m.s displacement a.

Further, we can generalize Eq. 3.3 by dividing the N particle system into M subsystems. Letting $N \to \infty$ while keeping M a finite integer, and defining $n = N/M$ we have

$$P_0(\sqrt{N}a; N) \approx \left[P_0(\sqrt{n}\ a; n)\right]^M = \epsilon(a)^N \tag{3.6}$$

where we define

$$\epsilon(a) := \left[P_0(\sqrt{n}\ a; n)\right]^{\frac{1}{n}} < 1 \quad . \tag{3.7}$$

In other words, if $N \gg 1$, the function $P_0(\sqrt{N}\ a; N)$ evaluated at fixed a decreases exponentially with N:

$$P_0(\sqrt{N}a; N) = e^{N \cdot \ln \epsilon(a)} \ . \tag{3.8}$$

This is the scaling property of $P_0(\sqrt{N}a; N)$ for a hard sphere crystal in the thermo-

dynamic limit.

3.2 Asymptotic formula for free energy

In this section, we derive a compact asymptotic formula (Eq. 3.19) for the free energy per particle in the thermodynamic limit. In this section and Sec. 3.3, our discussion is applicable to general crystalline solids, not just hard sphere solids. To emphasize the explicit N-dependence, let us write the partition function for a particular phase α as (see Eqs. 2.29 and 2.24)

$$Z_\alpha(N) = \frac{C_N P_0(r;N) s(r;N)}{p_{\mathcal{C}}^{\mathrm{eq}}(r;N)} \tag{3.9}$$

where $C_N = |\Omega|/\Lambda^{3N} = N^3|\omega|/\Lambda^{3N}$. As indicated earlier in the discussion of Eq. 2.29, $Z_\alpha(N)$ does not depend on the choice of the value of r.

In discussions of the thermodynamics limit, it is more convenient to replace r by $\sqrt{N}a$ (see Eq. 3.1) to address the N dependence explicitly. In such a limit, typical values of r scale as $\sqrt{N}$, but the typical value of intensive order parameter a does not diverge. As $N \to \infty$, the statistical fluctuations of a vanish and the value of a approaches a constant,

$$a_0 \equiv \lim_{N\to\infty} \frac{r_0}{\sqrt{N}} \tag{3.10}$$

where r_0 is the most likely value of r (see Fig. 2.4).

Using Eq. 3.9, the free energy per particle in the thermodynamic limit is:

$$\begin{aligned}\beta f_\alpha(\infty) = -\lim_{N\to\infty}\frac{1}{N}\Big[&\ln\left[C_N\cdot s(\sqrt{N}a;N)\right]\\&-\ln p_{\mathcal{C}}^{\mathrm{eq}}(\sqrt{N}a;N)+\ln P_0(\sqrt{N}a;N)\Big]\end{aligned} \tag{3.11}$$

which does not depend on the choice of the value a. In the following paragraphs, we evaluate the 3 terms on the right side of Eq. 3.11 at $a = a_0$, and we sum them to obtain the final result, Eq. 3.19.

For the first term in Eq. 3.11 we get

$$\lim_{N\to\infty} -\frac{1}{N}\ln\left[C_N\cdot s(\sqrt{N}a_0;N)\right] = -\frac{3}{2}\ln\frac{2\pi a_0^2}{3\Lambda^2}-\frac{3}{2} \quad . \tag{3.12}$$

by using Eqs. 2.26 and 2.27 and Stirling's formula for the Gamma function [38], $\ln\Gamma(z) = z\ln z - z + O(\ln z)$.

To evaluate the second term in Eq. 3.11 at $a = a_0$, let us define

$$x = a^2 = \frac{1}{N}\sum_{i=1}^{N}|\vec{r}_i|^2 \tag{3.13}$$

The probability density function of x is directly related to $p_{\mathcal{C}}^{\mathrm{eq}}(\sqrt{N}a;N)$ via a change of variables from $r = \sqrt{N}a$ to x:

$$p(x;N) = p_{\mathcal{C}}^{\mathrm{eq}}(\sqrt{N}a;N)\frac{\sqrt{N}}{2a} \tag{3.14}$$

For large N, the central limit theorem applies to the density function $p(x;N)$ that

can be approximated as a narrowly peaked Gaussian distribution of variance σ^2/N, where σ is system-dependent. The peak of the Gaussian is found at $x_0 = a_0^2$:

$$p(x_0; N \gg 1) \approx \sqrt{\frac{N}{2\pi\sigma^2}}\, e^{-N(x_0 - x_0)^2/2\sigma^2} = \sqrt{\frac{N}{2\pi\sigma^2}} \tag{3.15}$$

By combining Eqs. 3.14 and 3.15, we find that

$$\lim_{N\to\infty} p_{\mathcal{C}}^{\mathrm{eq}}(\sqrt{N}a_0; N) = \sqrt{\frac{2a_0^2}{\pi\sigma^2}} \quad . \tag{3.16}$$

Since both a_0 and σ are finite in the thermodynamic limit, we get

$$\lim_{N\to\infty} \frac{1}{N} \ln p_{\mathcal{C}}^{\mathrm{eq}}(\sqrt{N}a_0; N) = 0 \quad . \tag{3.17}$$

This result can be alternatively derived using the large deviation principle [39].

Finally for the third term in Eq. 3.11, we define the *rate function* of $P_0(\sqrt{N}a_0; N)$ as

$$\Phi(a_0) = -\lim_{N\to\infty} \frac{1}{N} \ln P_0(\sqrt{N}a_0; N) \ . \tag{3.18}$$

This rate function characterizes the rate of the exponential decay of $P_0(\sqrt{N}a_0; N)$ as $N \to \infty$. We have obtained a similar scaling relation specifically for hard sphere crystals (see Eq. 3.8). Let us assume for the moment that $\Phi(a_0)$ exists, i.e. it has a finite value.

Substituting Eqs. 3.12, 3.17 and 3.18 into Eq. 3.11 gives us

$$\beta f_\alpha(\infty) = \Phi(a_0) - \frac{3}{2}\ln\frac{2\pi a_0^2}{3\Lambda^2} - \frac{3}{2} \quad . \tag{3.19}$$

This result indicates that the free energy per particle in the thermodynamic limit is determined by two statistical quantities: the most likely r.m.s. displacement, a_0, and the rate function $\Phi(a_0)$[1]. While a_0 is relatively easy to estimate by direct MC simulation of a large crystal, the evaluation of the rate function $\Phi(a_0)$ is computationally expensive. Additionally, Eq. 3.19 indicates that $\Phi(a_0)$ must take a finite value as long as $f_\alpha(\infty)$ takes a finite value.

Eq. 3.19 can be easily generalized for D-dimensional solids:

$$\beta f_\alpha(\infty) = \Phi(a_0) - \frac{D}{2}\ln\frac{2\pi a_0^2}{D\Lambda^2} - \frac{D}{2} \quad . \tag{3.20}$$

3.3 Alternative derivation of Eq. 3.19

Instead of using the central limit theorem, we provide an alternative derivation of Eq. 3.19 by using large deviation theory [39].

Let us start from Eq. 3.11 and evaluate the right side term by term. The first term is evaluated as in Eq. 3.12. We use large deviation theory to evaluate the second term. As system size increases, the probability density function $p(x;N)$ (see Eq. 3.14) becomes narrowly peaked at $x_0 = a_0^2$. In the limit, $N \to \infty$, $p(x;N)$

[1] At Eq. 3.19, we have assumed that $\Phi(a_0)$ is finite. Now using Eq. 3.19, it is easy to show that as long as there exist a finite free energy per particle in the thermodynamic limit, then $\Phi(a_0)$ must take a finite value.

approximately becomes a delta function. In this limit, large deviation theory is a useful and convenient tool to describe $p(x;N)$ by its rate function (see Eq. 3.21). Note that variable x is an arithmetic mean of N random variables and these random variables are the square displacements of each particle from its lattice site. In the thermodynamic limit, we can assume the square displacements to be independent from each other. Under this assumption, it is useful to consider the following limit [39]:

$$\lim_{N\to\infty} -\frac{1}{N}\log p(x;N) = I(x) \tag{3.21}$$

where $I(x)$ is called a rate function. From the properties of the rate functions in large deviation theory, $I(x)$ is a convex function with minimum 0 that is taken at the peak of the probability density function ($x_0 := a_0^2$). Using Eq. 3.14, it is easy to verify that

$$\lim_{N\to\infty} -\frac{1}{N}\log p_{\mathcal{C}}^{\mathrm{eq}}(\sqrt{N}a;N) = \lim_{N\to\infty} -\frac{1}{N}\log p(x;N) = I(a^2) \quad . \tag{3.22}$$

For the third term of Eq. 3.11, let us assume that the rate function $\Phi(a)$ defined in Eq. 3.18 exists for all values of a.

$$\Phi(a) = -\lim_{N\to\infty}\frac{1}{N}\ln P_0(\sqrt{N}a;N) \tag{3.23}$$

Then by summing the three terms (see Eqs. 3.12, 3.22 and 3.23), the asymp-

totic free energy becomes

$$\beta f_\alpha(\infty) = \Phi(a) - I(a^2) - \frac{3}{2}\ln\frac{2\pi a^2}{3\Lambda^2} - \frac{3}{2} \quad . \tag{3.24}$$

where, $f_\alpha(\infty)$ is independent of the choice of a. Notice that using the property that $I(a_0^2) = 0$, and choosing $a = a_0$, Eq. 3.24 reduces to Eq. 3.11.

Chapter 4: Separation of Translational and Internal Degrees of Freedom.

4.1 Motivation

Computational statistical physics provides a useful set of tools to estimate the absolute free energies of different macroscopic systems (e.g. crystalline solids, fluids and so on). This estimation however is an indirect extrapolation. In practice, numerical simulations can only be performed on limited numbers of particles. These numbers are much smaller than that are found in macroscopic systems (usually larger than 10^{20}). To estimate the free energy per particle of the macroscopic system, a series of simulations for the same system of different sizes (i.e. number of particles) are performed. As the system size increases, if the intensive free energies per particle converge to a stationary value, then this value is believed to be the same as the intensive free energy in the thermodynamic limit (of the same macroscopic system). Various numerical methods that was developed in the past decades [11–23] as well as the new method described in Chapter 2 provide good estimates of the free energies of such microscopic systems.

Microscopic systems are particularly sensitive to the boundary condition (shape

or type). In extrapolating microscopic free energies, the boundary effects (or finite size effect) need to be reduced for better convergence. A widely accepted approach is to apply periodic boundary conditions on the simulation box, where the particles interact with the particles from the opposite side of the box, instead of an artificial hard wall. However, for simulations adopting such boundary conditions, there are two subtleties that need to be considered carefully. Firstly, the definition of the free energy (or the partition function) needs to accommodate its corresponding boundary conditions, which should be reflected in either the form of the Hamiltonian and/or the domain of the phase space. Secondly, unlike hard boundary conditions, periodic boundary conditions allow for translations of the entire system. In numerical simulations, these translational degrees of freedom are usually held fixed, but a naive artificial constraint on the translation can lead to nonnegligible errors, as reported in [1, 10, 31, 40].

In Chapter 2, we defined the partition function for a crystal within periodic boundary conditions (see Sec. 2.2.1) and introduced a change of coordinates that separates the translational and internal degrees of freedom (see Sec. 2.2.2). In this chapter we provide a clear definition of the free energy that accommodates periodic boundary conditions and discuss several frameworks to separate and fix the translational and internal degrees of freedom for evaluation of the free energy. [1].

[1]Here we avoid using the term "center of mass" to characterize the translational movement for the following reason: center of mass is no longer naturally defined for systems with periodic boundary conditions. In face, we can use any description of the external translation, as long as the translational and internal degrees of freedoms are separated.

4.2 Definitions and notations

The following discussion applies to systems at fixed temperature T that consist of a fixed number of particles (interacting or noninteracting) that are confined within a periodic box with a fixed volume. We denote the mass of the i-th particle by m_i, and we do not presume the indistinguishability of the particles. Let us denote the D-dimensional simulation box by $\mathbf{\Omega}$, upon which periodic boundary conditions are applied. We denote the positions of the N particles by $\vec{q}_i$'s and the momenta by $\vec{p}_i$'s. The microstate of the system is denoted by

$$\Gamma = (\mathbf{p},\ \mathbf{q}) = (\vec{p}_1,\ \cdots,\ \vec{p}_N,\ \vec{q}_1,\ \cdots,\ \vec{q}_N) \tag{4.1}$$

and includes the momenta and positions of all of the particles. Note that the momenta space is unbounded but the configuration space is bounded due to the rectangular simulation box:

$$\mathbf{p} \in \mathbb{R}^{ND} \tag{4.2}$$

$$\mathbf{q} \in \mathbf{\Omega}^N \subset \mathbb{R}^{ND} \tag{4.3}$$

We assume that the geometric center of the box is located at the origin and its edges are oriented along the directions of basis vectors $(\hat{x},\ \hat{y},\ \cdots)$. The Hamiltonian is

assumed to be the sum of a kinetic energy and a potential energy,

$$H(\Gamma) = E_k(\mathbf{p}) + U(\mathbf{q}) \tag{4.4}$$

This function maps any microstate Γ from the phase space $\mathbb{R}^{ND} \otimes \mathbf{\Omega}^N$ to a real number – the total energy of the microstate. Without losing generality, we restrict the potential energy to be the sum of all pairwise interactions between particles, and thus it is invariant under translation of the entire system[2]. We also require that the particle-particle interactions vanish at a cutoff range (distance), which must be smaller than half of the shortest edge-length of the simulation box $\mathbf{\Omega}$. Such cutoffs prevent a particle from interacting with itself, or with multiple period images of another particle. We specifically write the potential energy as the sum of particle-particle interactions:

$$U(\vec{q}_1, \ \cdots, \ \vec{q}_N) = \frac{1}{2} \sum_{i \neq j} U_{ij} \left(\min_{Z_{d\times d}} \| \vec{q}_i - \vec{q}_j + Z_{d\times d} \cdot \vec{\omega} \| \right) \tag{4.5}$$

where $Z_{D\times D}$ is an arbitrary[3] D by D integer-element diagonal matrix, $\vec{\omega}$ is a D-dimensional vector whose elements are the edge lengths of the simulation box $\mathbf{\Omega}$, and $\|\cdot\|$ is the Euclidean norm (distance). Note that the argument of the pair potential $U_{ij}(\cdot)$ provides the minimum distance between $\vec{q}_i$ and the infinite periodic images of $\vec{q}_j$. The cut-off range of the pair potential is required to be less than half of

[2]It is not necessary to restrict the interaction by two-body interactions. Our conclusion is true as long as the system's potential energy is translationally invariant, i.e. no external potential.

[3]In fact, since $\vec{q}_i \in \mathbf{\Omega}$, and $\vec{q}_j \in \mathbf{\Omega}$, it is sufficient to choose the diagonal elements of $Z_{d\times d}$ from $\{0, \ 1, \ -1\}$, instead of the whole set of integers.

the smallest component of $\vec{\omega}$. We sketch the pair interactions as directed arrows in Fig. 4.1.

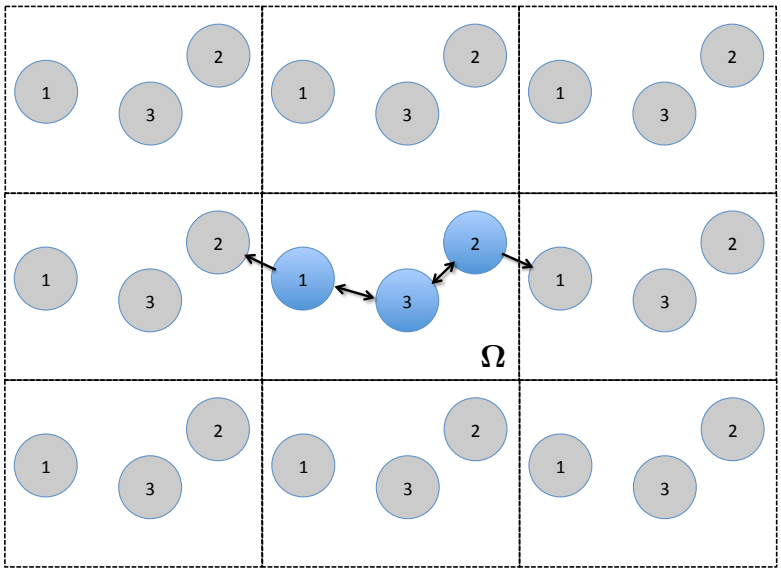


Figure 4.1: An illustration of the interactions between particles with periodic boundary conditions. Periodic images of particles are gray and the original images of particles in the simulation box are blue. Taking particle #1 as an example, it interacts with the periodic image of particle #2 on the left, and also interacts with the original image of particle #3. The six interaction terms in Eq. 4.5 are shown as directed arrows.

At this point, having defined both the Hamiltonian and its domain, we write the partition function for this finite system as

$$\begin{aligned} Z &= \frac{1}{c_N h^{ND}} \int_{\mathbb{R}^{ND}} \mathrm{d}\mathbf{p}\; e^{-\beta E_k(\mathbf{p})} \cdot \int_{\mathbf{\Omega}^N} \mathrm{d}\mathbf{q}\; e^{-\beta U(\mathbf{q})} \\ &= \frac{1}{c_N h^{ND}}\; P \cdot Q \end{aligned} \tag{4.6}$$

where $c_N = N!$ (or 1) if the N particles are indistinguishable (distinguishable). The momentum integral P can be calculated analytically:

$$P = \prod_{i=1}^{N} \left(\frac{\beta}{2\pi m_i} \right)^{\frac{D}{2}} . \tag{4.7}$$

In the following discussion, we explore different coordinate systems that separate the translational and internal contributions to the configurational integral

$$Q = \int_{\mathbf{\Omega}} \mathrm{d}\vec{q}_1 \cdots \int_{\mathbf{\Omega}} \mathrm{d}\vec{q}_N \; e^{-\beta U(\vec{q}_1, \cdots, \vec{q}_N)} \quad . \tag{4.8}$$

4.3 Separation of translational and internal degrees of freedom

In the rest of the chapter, we use an example consisting of 3 particles ($N = 3$) within a one-dimensional space ($D = 1$), with periodic boundary conditions, and sketch the configuration space and the separation of the translational degrees of freedom. This is the most complicated system whose configuration space can be visualized in 3 dimensions. But our derivations and results are general for any finite N and D. The example system is illustrated in Fig. 4.2.

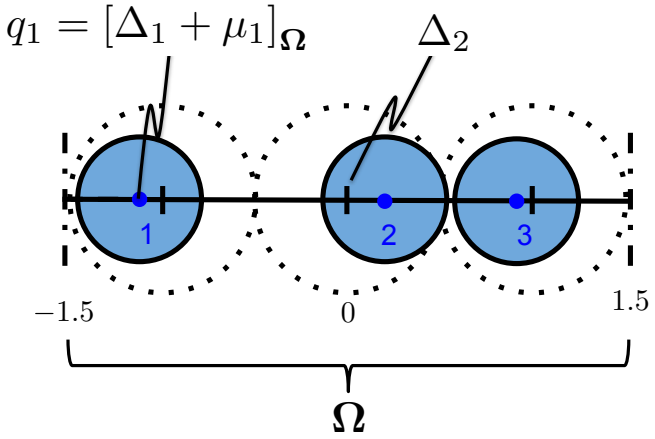


Figure 4.2: Three hard particles (shaded circles) are confined in a one-dimensional periodic boundary container. The periodic boundaries are illustrated by the vertical dashed lines at -1.5 and 1.5. Here $\mathbf{\Omega} = [-1.5, 1.5)$. The perfect crystal configuration corresponds to a close packed configuration, where each particle has the maximum size that is equal to 1. We define $\Delta_i := i - 2$ and the actual positions of the particles are denoted by q_i.

First, let us associate each particle with a reference position $\vec{\Delta}_i \in \mathbf{\Omega}$. We

require that the geometrical center of the set of reference positions is located at the origin:

$$\sum_{i=1}^{N} \vec{\Delta}_i = \vec{0} \quad . \tag{4.9}$$

For fluid systems, one can choose $\vec{\Delta}_i = \vec{0}$ for all the particles. However, in studying crystalline solids, it makes more physical sense to choose $\vec{\Delta}_i$'s using the positions of the corresponding perfect crystal lattice. In other words, this set of reference positions defines a perfect lattice contained in $\mathbf{\Omega}$, and the geometrical center of the lattice is located at the origin of $\mathbf{\Omega}$. In the following discussion we assume our system is a crystalline solid with a unique stable phase. But this discussion is immediately applicable to fluid systems and solids with multiple phases (e.g. FCC, HCP and so on).

Using the set of reference positions, let us consider the following change of coordinates from $\mathbf{q}$ to $\boldsymbol{\mu}$:

$$\vec{\mu}_i = \left[\vec{q}_i - \vec{\Delta}_i\right]_{\mathbf{\Omega}} \tag{4.10}$$

where the operation $[\cdot]_{\mathbf{\Omega}}$ takes any D-dimensional vector back into the periodic box $\mathbf{\Omega}$. Thus, $\vec{\mu}_i \in \mathbf{\Omega}$ for all i's from 1 to N. It is easy to verify that the Jacobian of this change of coordinate is 1.

Then, we can write

$$\begin{aligned} Q &= \int_{\mathbf{\Omega}_1} \mathrm{d}\vec{\mu}_1 \cdots \int_{\mathbf{\Omega}_N} \mathrm{d}\vec{\mu}_N \; e^{-\beta V(\vec{\mu}_1, \cdots, \vec{\mu}_N)} \\ &= \int_{\mathbf{\Omega}} \mathrm{d}\vec{\mu}_1 \cdots \int_{\mathbf{\Omega}} \mathrm{d}\vec{\mu}_N \; e^{-\beta V(\vec{\mu}_1, \cdots, \vec{\mu}_N)} \end{aligned} \tag{4.11}$$

where the $\mathbf{\Omega}_i$'s are $\mathbf{\Omega}$ shifted by $\vec{\Delta}_i$. The second equality is justified by the periodic boundary conditions and the way the potential energy is defined. At this point, we have a new variable $\boldsymbol{\mu} \in \mathbf{\Omega}^N$ to describe the configuration. The bounded configuration space is $\mathbf{\Omega}^N$ in the $\boldsymbol{\mu}$-space. Note that the potential energy defined in the new coordinate system is $V\big(\boldsymbol{\mu}\,(\mathbf{q})\,\big) = U(\mathbf{q})$. More specifically,

$$V(\vec{\mu}_1,\ \cdots,\ \vec{\mu}_N) = \frac{1}{2}\sum_{i\neq j} U_{ij}\left(\min_{Z_{d\times d}} \|\vec{\mu}_i + \vec{\Delta}_i - \vec{\mu}_j - \vec{\Delta}_j + Z_{d\times d}\cdot\vec{\omega}\|\right) \tag{4.12}$$

We illustrate the $\boldsymbol{\mu}$-space using the 3-particle example and also illustrate the set of allowed configurations within $\mathbf{\Omega}^3$ in Fig. 4.3.

To specify the translational degrees of freedom, we define a mapping $\vec{c}_0(\cdot)$ that maps any configuration $\boldsymbol{\mu} \in \mathbf{\Omega}^N$ to a D-dimensional vector:

$$\vec{c}_0(\boldsymbol{\mu}) = \frac{1}{N}\left[\sum_{i=1}^{N}\vec{\mu}_i\right]_{\mathbf{\Omega}} \in \mathbb{T} \subset \mathbb{R}^D \tag{4.13}$$

Here the set of configurations is mapped to $\mathbb{T}$, a D-dimensional region inside $\mathbf{\Omega}$ that is centered at the origin. Each edge length of $\mathbb{T}$ is $1/N$ of the length of the corresponding edge for $\mathbf{\Omega}$, thus

$$|\mathbb{T}| = \frac{1}{N^D}|\mathbf{\Omega}| \tag{4.14}$$

In the 3-particle example (see Fig. 4.2), the mapping $\vec{c}_0(\boldsymbol{\mu})$ is 1-dimensional, and its range is $\mathbb{T} = [-1.5/N, 1.5/N) = [-0.5, 0.5)$, which is easy to verify.

In the following discussion, we will use $\vec{c}_0$ to describes the translational degrees

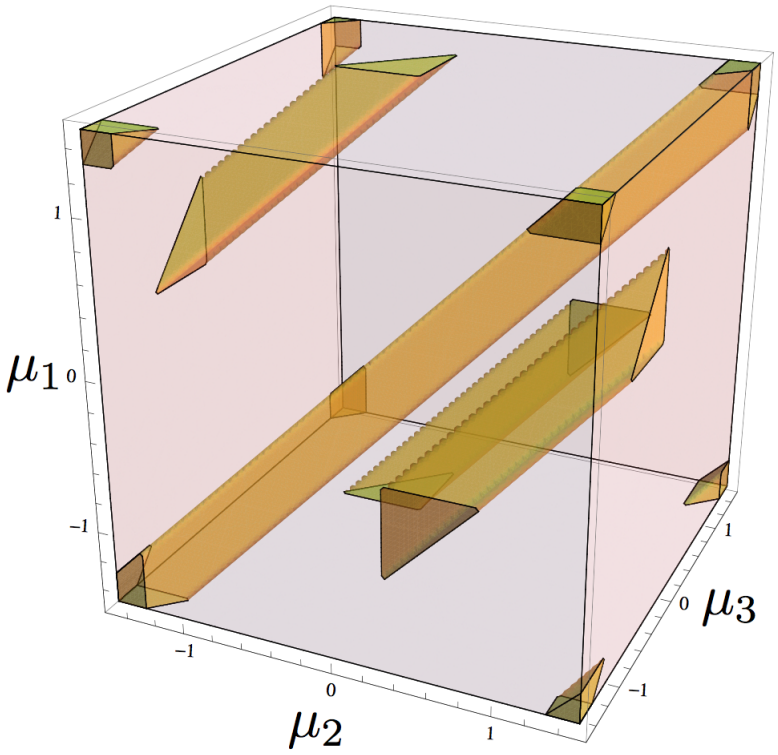


Figure 4.3: Using the example system (see Fig. 4.2), we sketch the configuration space $\boldsymbol{\mu} \in \boldsymbol{\Omega}^3 = [-1.5, 1.5)^3$ as a pink cube. By assuming the diameter of the one-dimensional hard sphere particles to be 0.9, we plot the set of energetically allowed configurations, shown as 10 disconnected regions colored in yellow. Note that there are 6 possible arrangements of the 3 particles. If we consider the $\boldsymbol{\mu}$-space as a 3 dimensional torus (i.e. using periodic boundary conditions), then the 10 disconnected sets of allowed configurations become two simply connected sets. One set of allowed configurations contains 3 different arrangements of particles –"1,2,3", "2,3,1", and "3,1,2"; while another set contains the rest 3 possible arrangements of the particles – "1,3,2", "3,2,1", and "2,1,3".

of freedom, and find the set of coordinates for the remaining (internal) degrees of freedom. Here we will only introduce two out of many different ways of describing the internal degrees of freedom.

4.3.1 Change of coordinates #1

Here we introduce a new set of coordinates $\boldsymbol{\nu}$. We will define this change of coordinates through a change of the sets of basis vectors. First, let us decompose the configuration in the old coordinates ($\boldsymbol{\mu}$-space) in terms of the (ND) basis vectors

$$\left(\mu_1^{(1)},\ \mu_1^{(2)},\ \cdots,\ \mu_1^{(D)},\ \mu_2^{(1)},\ \cdots,\ \mu_N^{(D)}\right)_{\boldsymbol{\mu}} = \sum_{i=1}^{N}\sum_{l=1}^{D} \mu_i^{(l)} \cdot \hat{\mu}_i^{(l)} \tag{4.15}$$

where the (ND) basis vectors $\hat{\mu}_i^{(l)}$ are orthonormal basis vectors corresponding to the old set of coordinates. A configuration can be represented by a weighted sum of another set of basis vectors shown as below:

$$\left(c_0^{(1)},\ c_0^{(2)},\ \cdots,\ c_0^{(D)},\ \nu_2^{(1)},\ \cdots,\nu_N^{(D)}\right)_{\boldsymbol{\nu}} = \sum_{l=1}^{D} c_0^{(l)} \cdot \hat{c}_0^{(l)} + \sum_{i=2}^{N}\sum_{l=1}^{D} \nu_i^{(l)} \cdot \hat{\nu}_i^{(l)} \tag{4.16}$$

where the new set of basis vectors $(\hat{c}_0^{(1)}, \cdots, \hat{\nu}_N^{(D)})$ define a new coordinate system. In the new coordinate system, we define the first D basis vectors in the following way so that they correspond to the translational (external) degrees of freedom of the whole system:

$$\hat{c}_0^{(l)} := \frac{1}{N}\sum_{i=1}^{N} \hat{\mu}_i^{(l)} \quad \text{for } l = 1,\ \cdots,\ D. \tag{4.17}$$

Regarding the internal degrees of freedom, we do not specifically define every $\hat{\nu}_i^{(l)}$ for i from 2 to N and l from 1 to D. It is adequate to simply require that these $N(D-1)$ basis vectors be orthonormal to each other,

$$\hat{\nu}_i^{(l)} \cdot \hat{\nu}_{i'}^{(l')} = \delta_{i,i'} \cdot \delta_{l,l'} \tag{4.18}$$

and orthogonal to the $\hat{c}_0^{(l)}$'s – the basis vectors for the translational degrees of freedom

$$\hat{\nu}_i^{(l)} \cdot \hat{c}_0^{(l')} = 0 \ , \quad \forall \ l, l' \in \{1,\ 2,\ , \cdots, d\} \quad . \tag{4.19}$$

We illustrate the new set of basis vectors using the 3-particle example and plot the basis vectors in Fig. 4.4.

For such a change of coordinates, the Jacobian can be calculated by comparing the old and new basis vectors:

$$|J_1| = \left| \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\nu}} \right| = \frac{|\hat{\mu}_1^{(1)}| \cdots |\hat{\mu}_N^{(D)}|}{|\hat{c}_0^{(1)}| \cdots |\hat{\nu}_N^{(D)}|} = N^{D/2} \tag{4.20}$$

Note that because of Eq. 4.17, $\hat{c}_0^{(l)}$, the first D basis vectors in the new basis are not normalized, but have magnitudes $N^{-1/2}$.

Finally, we separate the contribution from external and internal degrees of

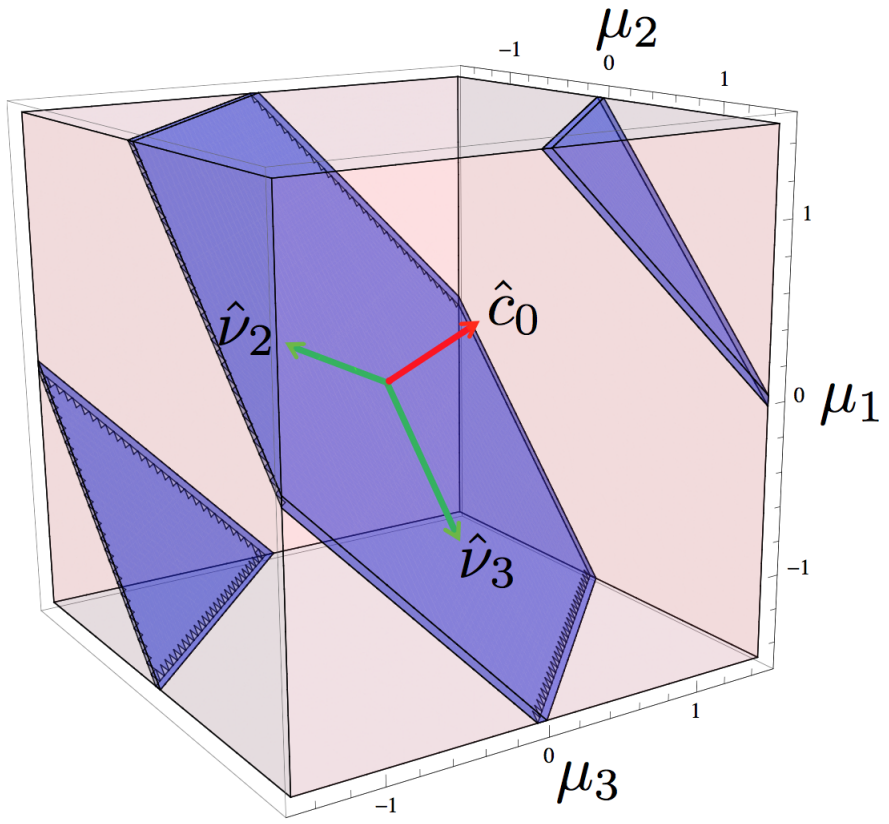


Figure 4.4: As an example, we sketch the 3 basis vectors corresponding to the new set of coordinates ($\boldsymbol{\nu}$). The basis vector $\hat{c}_0$ for the external degrees of freedom (see Eq. 4.13) is shown as a red arrow. The basis vectors for internal degrees of freedom are shown as the two green arrows. The choice of orthonormal vectors $\hat{\nu}_2$ and $\hat{\nu}_3$ are not unique, as long as they are perpendicular to $\hat{c}_0$. The magnitude of $\hat{c}_0$ is $N^{-1/2} = 1/\sqrt{3}$, and the magnitudes of $\hat{\nu}_2$, and $\hat{\nu}_3$ are both unity. The blue planes (one hexagon and two triangles) are the set of configurations – $\boldsymbol{\mu}$ that satisfy $\vec{c}_0(\boldsymbol{\mu}) = \vec{0}$.

freedom to Q:

$$\begin{aligned} Q =& |J_1| \cdot \int_{\mathbb{T}} \mathrm{d}\vec{c}_0 \int \cdots \int_{\mathbb{P}(\vec{c}_0)} \mathrm{d}\vec{\nu}_2 \cdots \mathrm{d}\vec{\nu}_N \; e^{-\beta E(\vec{c}_0, \vec{\nu}_2, \cdots, \vec{\nu}_N)} \\ &= N^{D/2} \cdot \int_{\mathbb{T}} \mathrm{d}\vec{c}_0 \; Q_{\text{int}}(\vec{c}_0) \\ &= N^{D/2} \cdot Q_{\text{int1}} \cdot \int_{\mathbb{T}} \mathrm{d}\vec{c}_0 \\ &= N^{D/2} \cdot Q_{\text{int1}} \cdot |\mathbb{T}| \end{aligned} \tag{4.21}$$

where we denote the potential energy defined in the new coordinate system as $E(\boldsymbol{\nu})$, and we denote the integral domain for $\vec{c}_0$ as $\mathbb{T}$. The integral domain for $(\vec{\nu}_2 \; \cdots \; \vec{\nu}_N)$ is a set of bounded hyperplanes (or *hyper-plates*) $\mathbb{P}$ that are perpendicular to the external degrees of freedom, and their shape (boundary) is a function of $\vec{c}_0$. (We sketch $\mathbb{P}(\vec{c}_0)$ for $\vec{c}_0 = \vec{0}$ as blue planes in both Fig. 4.3 and Fig. 4.4. Note that the energy function $E(\boldsymbol{\nu})$ defines an energy landscape on any plane $\mathbb{P}(\vec{c}_0)$ defined by $\vec{c}_0 \in \mathbb{T}$, and we sketch the set of allowed configurations on $\mathbb{P}(\vec{c} = \vec{0})$ in Fig. 4.5.

In the third equality in Eq. 4.21, we assert that the integral over the internal degrees of freedom is independent of the choice of the translational vector $\vec{c}_o$, which is a direct effect of the periodic boundary condition. In other words,

$$\begin{aligned} Q_{\text{int1}} :=& \int \cdots \int_{\mathbb{P}(\vec{c}_0)} \mathrm{d}\vec{\nu}_2 \cdots \mathrm{d}\vec{\nu}_N \; e^{-\beta E(\vec{c}_0, \vec{\nu}_2, \cdots, \vec{\nu}_N)} \\ =& \int \cdots \int_{\mathbb{P}(\vec{c}_0)} \mathrm{d}\vec{\nu}_2 \cdots \mathrm{d}\vec{\nu}_N \; e^{-\beta E(\vec{\nu}_2, \cdots, \vec{\nu}_N)} \\ =& \text{ A constant independent of } \vec{c}_0 \end{aligned} \tag{4.22}$$

With this change coordinates, we separate the external degrees of freedom

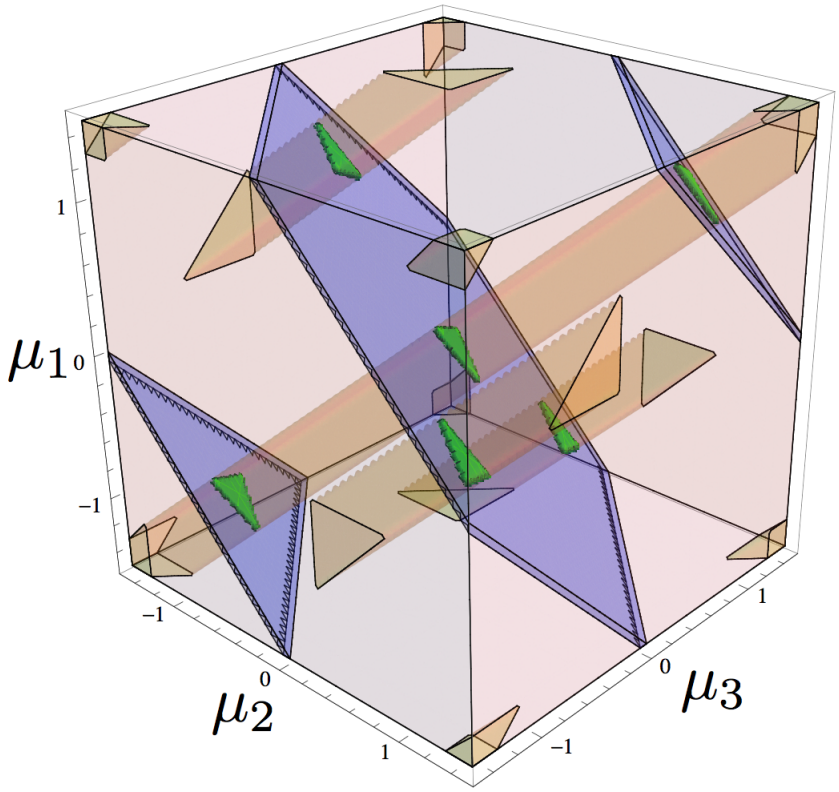


Figure 4.5: Using the example of a 3-particle system, we sketch $\mathbb{P}(0)$, the set of configurations satisfying $\vec{c}_0 = 0$ as blue plates. The set of allowed configurations intersects with these plates, and the intersections are shown as 6 green triangles. Here Q_{int1} in this case (hard sphere potential) is simply the total measure (area) of the 6 green triangular pieces. Note that the 6 triangular pieces have the same measure, and they are related by particle-particle permutation. Each triangular piece contains one of the 3! arrangements of the 3 particles.

$(\vec{c}_0)$ and the internal degrees of freedom $(\vec{\nu}_2, \cdots, \vec{\nu}_N)$. The advantage of this coordinate system compared with the traditionally used center of mass is that as long as the system is translationally invariant, the translational degrees of freedom are sampled uniformly at equilibrium. Thus, the integral over the translational degrees of freedom is simply the measure of the domain $\mathbb{T}$, which is equal to the measure of the simulation box divided by N^D (Eq. 4.14). To calculate the partition function, we need to compute Q_{int1}, the integral of the Boltzmann factor over the internal degrees of freedom. Since the system is translationally invariant, or independent of the translational degrees of freedom, $\vec{c}_0$, we can choose any fixed value of $\vec{c}_0 \in \mathbb{T}$, and numerically evaluate Q_{int1}.

$$Q = |J_1| \cdot |\mathbb{T}| \cdot Q_{\text{int1}} = \frac{|\mathbf{\Omega}|}{N^{d/2}} \cdot Q_{\text{int1}} \tag{4.23}$$

4.3.2 Change of coordinates #2

Alternatively, we can rewrite Eq. 4.6 by introducing another change of coordinates from $\mathbf{q}$ to $\boldsymbol{\gamma}$, where

$$\boldsymbol{\gamma} = (\vec{c}_0, \ \vec{\mu}_2, \ \cdots, \ \vec{\mu}_N), \tag{4.24}$$

where $\vec{c}_0$ is defined by Eq. 4.13 and the $\vec{\mu}_i$'s (for $i > 1$) are defined by Eq. 4.10. The position of the first particle appears to be missing in these new coordinates. However, one can uniquely determine $\mathbf{q}$, the set of positions of all particles from any

$\boldsymbol{\gamma}$ by the following inverse map:

$$\begin{aligned}\vec{q}_1 &= \left[N\cdot\vec{c}_0 + \vec{\Delta}_1 - \vec{\mu}_2 - \vec{\mu}_3 - \cdots - \vec{\mu}_N\right]_{\boldsymbol{\Omega}} \\ \vec{q}_i &= \left[\vec{\Delta}_i + \vec{\mu}_i\right]_{\boldsymbol{\Omega}} \quad , \quad i > 1.\end{aligned} \tag{4.25}$$

To obtain Eqs. 4.25, we first obtain $\vec{q}_i$ for $i > 1$ directly from Eq. 4.10 and the fact that $\vec{q}_i \in \boldsymbol{\Omega}$. Similarly we can write $\vec{\Delta}_i = [\vec{q}_i - \vec{\mu}_i]_{\boldsymbol{\Omega}}$. To obtain $\vec{q}_1$ in Eq. 4.25, we need to rewrite Eq. 4.9 and as

$$\begin{aligned}\vec{0} &= \vec{\Delta}_1 + \cdots + \vec{\Delta}_N \\ &= [\vec{q}_1 - \vec{\mu}_1]_{\boldsymbol{\Omega}} + \cdots + [\vec{q}_N - \vec{\mu}_N]_{\boldsymbol{\Omega}} \\ &= \left[\sum_{i=1}^{N}\vec{q}_i - \sum_{i=1}^{N}\vec{\mu}_i\right]_{\boldsymbol{\Omega}} \\ &= \left[\sum_{i=1}^{N}\vec{q}_i - \left[\sum_{i=1}^{N}\vec{\mu}_i\right]_{\boldsymbol{\Omega}}\right]_{\boldsymbol{\Omega}} \\ &= \left[\vec{q}_1 + \sum_{i=2}^{N}\vec{q}_i - N\cdot\vec{c}_0\right]_{\boldsymbol{\Omega}} \\ &= \left[\vec{q}_1 + \sum_{i=2}^{N}\left[\vec{\Delta}_i + \vec{\mu}_i\right]_{\boldsymbol{\Omega}} - N\cdot\vec{c}_0\right]_{\boldsymbol{\Omega}} \\ &= \left[\vec{q}_1 - \vec{\Delta}_1 + \sum_{i=2}^{N}\vec{\mu}_i - N\cdot\vec{c}_0\right]_{\boldsymbol{\Omega}}\end{aligned} \tag{4.26}$$

where we used the following property of the periodic boundary condition operator $[\cdot]_{\boldsymbol{\Omega}}$:

$$\left[[\vec{a}]_{\boldsymbol{\Omega}} + [\vec{b}]_{\boldsymbol{\Omega}}\right]_{\boldsymbol{\Omega}} = \left[\vec{a} + \vec{b}\right]_{\boldsymbol{\Omega}} \quad . \tag{4.27}$$

With the change of the coordinates, we can write

$$\begin{aligned} Q &= |J_2| \cdot \int_{\mathbb{T}} \mathrm{d}\vec{c}_0 \int_{\mathbf{\Omega}} \mathrm{d}\vec{\mu}_2 \ \cdots \int_{\mathbf{\Omega}} \mathrm{d}\vec{\mu}_N \ e^{-\beta \mathbf{E}(\vec{c}_0,\ \vec{\mu}_2,\ \cdots,\ \vec{\mu}_N)} \\ &= |J_2| \cdot \int_{\mathbb{T}} \mathrm{d}\vec{c}_0 \ Q_{\mathrm{int2}}(\vec{c}_0) \\ &= |J_2| \cdot Q_{\mathrm{int2}} \cdot \int_{\mathbb{T}} \mathrm{d}\vec{c}_0 \\ &= |J_2| \cdot Q_{\mathrm{int2}} \cdot |\mathbb{T}| \end{aligned} \tag{4.28}$$

where the integral domain for the internal degrees of freedom is the hyper-box $\mathbf{\Omega}^{D-1}$, independent of the choice of $\vec{c}_0$. This change of coordinates is a one to one mapping, and we need to find its Jacobian. Let us consider the change of coordinates:

$$\vec{c}_0 = \frac{1}{N}\left[\sum_{i=1}^{N} \vec{\mu}_i\right]_{\mathbf{\Omega}} = \frac{1}{N}\left[\sum_{i=1}^{N} \left[\vec{q}_i - \vec{\Delta}_i\right]_{\mathbf{\Omega}}\right]_{\mathbf{\Omega}} = \frac{1}{N}\left[\sum_{i=1}^{N} \vec{q}_i\right]_{\mathbf{\Omega}} \tag{4.29}$$

which takes advantage of Eq. 4.9 and the property of the periodic boundary operation. We can also change $\boldsymbol{\gamma}$ back to $\mathbf{q}$ by using Eq. 4.25.

Then to calculate the Jacobian $|J_2|$, we first write the matrix $\partial\mathbf{q}/\partial\boldsymbol{\gamma}$ in block-diagonal form, by separately grouping the contributions from the D dimensions:

$$\frac{\partial \mathbf{q}}{\partial \boldsymbol{\gamma}} = \begin{pmatrix} \mathbf{J}^{(1)} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{J}^{(2)} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{J}^{(D)} \end{pmatrix}_{ND \times ND} . \tag{4.30}$$

From Eq. 4.25 we have

$$\mathbf{J}^{(l)} = \begin{pmatrix} N & -1 & -1 & -1 & \cdots & -1 \\ 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 \end{pmatrix}_{N\times N} , \; l = 1,\; 2,\; \cdots ,\; D. \tag{4.31}$$

The determinant of this matrix is N, as computed using the expansion of minors. Since $\partial\mathbf{q}/\partial\boldsymbol{\gamma}$ is block-diagonal, we obtain

$$|J_2| = \left|\frac{\partial \mathbf{q}}{\partial \boldsymbol{\gamma}}\right| = \prod_{l=1}^{D} |\mathbf{J}^{(l)}| = N^D \quad . \tag{4.32}$$

For this set of coordinates we can write

$$Q = |J_2| \cdot |\mathbb{T}| \cdot Q_{\text{int2}} = |\boldsymbol{\Omega}| \cdot Q_{\text{int2}} \tag{4.33}$$

where

$$Q_{\text{int2}} = \int_{\boldsymbol{\Omega}} \mathrm{d}\vec{\mu}_2 \; \cdots \int_{\boldsymbol{\Omega}} \mathrm{d}\vec{\mu}_N \; e^{-\beta \mathbf{E}(\vec{c}_0,\, \vec{\mu}_2,\, \cdots,\, \vec{\mu}_N)}. \tag{4.34}$$

4.4 Discussion of other methods

4.4.1 Reduction to the method used in Chapter 2

As discussed above, the value of Q_{int2} is independent of the choice of $\vec{c}_0 \in \mathbb{T}$. Here, let us choose $\vec{c}_0 = \vec{0}$ and we have $\vec{q}_i^{\,\text{lat}} = \Delta_i$. As a result $\vec{r}_{i\neq 1} = \vec{\mu}_{i\neq 1}$, and we can write

$$\begin{aligned} Q_{\text{int2}} &= \int_{\mathbf{\Omega}} \mathrm{d}\vec{\mu}_2 \ \cdots \int_{\mathbf{\Omega}} \mathrm{d}\vec{\mu}_N \ e^{-\beta \mathbf{E}(\vec{c}_0,\ \vec{\mu}_2,\ \cdots,\ \vec{\mu}_N)} \\ &= \int_{\mathbf{\Omega}} \mathrm{d}\vec{r}_2 \ \cdots \int_{\mathbf{\Omega}} \mathrm{d}\vec{r}_N \ e^{-\beta \mathbf{E}(\vec{0},\ \vec{r}_2,\ \cdots,\ \vec{r}_N)} \\ &= \int_{\mathbb{R}^D} \mathrm{d}\vec{r}_1 \int_{\mathbf{\Omega}} \mathrm{d}\vec{r}_2 \ \cdots \int_{\mathbf{\Omega}} \mathrm{d}\vec{r}_N \ e^{-\beta \mathbf{E}(\vec{0},\ \vec{r}_2,\ \cdots,\ \vec{r}_N)} \cdot \delta^D(\vec{r}_1 + \vec{r}_2 + \cdots + \vec{r}_N) \end{aligned} \tag{4.35}$$

which is equal to $\int' \mathrm{d}\mathbf{r} \ e^{-\beta V}$ (see Eq. 2.19 in Chapter 2). Then Eq. 4.33 reduces to the result in Chapter 2, where we had

$$Q = |\mathbf{\Omega}| \cdot \int' \mathrm{d}\mathbf{r} \ e^{-\beta V} \quad . \tag{4.36}$$

4.4.2 Method due to Vega and Noya

In 2007, Vega and Noya [17] proposed a different method to separate the translational and internal degrees of freedom. Instead of collectively defining translational degrees of freedom, they chose the position of one (the first) particle to carry the translational degrees of freedom. In their method, the positions of the rest of particles are represented by their relative position with respect to the chosen particle.

Their method can be formally stated by a change of coordinates from $\mathbf{q}$ to

$$\boldsymbol{\eta} = (\vec{\eta}_1,\ \vec{\eta}_2,\ \cdots,\ \vec{\eta}_N) \tag{4.37}$$

where $\vec{\eta}_1 = \vec{q}_1 \in \mathbf{\Omega}$, and for $i = 2, \cdots, N$,

$$\vec{\eta}_i = [\vec{q}_i - \vec{q}_1]_{\mathbf{\Omega}} \tag{4.38}$$

In the new coordinate system, the configuration integral Q is written as,

$$\begin{aligned} Q &= |J_3| \cdot \int_{\mathbf{\Omega}} \mathrm{d}\vec{q}_1 \int_{\mathbf{\Omega}} \mathrm{d}\vec{\eta}_2 \ \cdots \int_{\mathbf{\Omega}} \mathrm{d}\vec{\eta}_N \ e^{-\beta E_{\mathrm{Vega}}(\vec{q}_1,\ \vec{\eta}_2,\ \cdots,\ \vec{\eta}_N)} \\ &= |J_3| \cdot \int_{\mathbf{\Omega}} \mathrm{d}\vec{q}_1 \int_{\mathbf{\Omega}} \mathrm{d}\vec{\eta}_2 \ \cdots \int_{\mathbf{\Omega}} \mathrm{d}\vec{\eta}_N \ e^{-\beta E_{\mathrm{Vega}}(\vec{\eta}_2,\ \cdots,\ \vec{\eta}_N)} \\ &= 1 \cdot |\mathbf{\Omega}| \cdot \int_{\mathbf{\Omega}} \mathrm{d}\vec{\eta}_2 \ \cdots \int_{\mathbf{\Omega}} \mathrm{d}\vec{\eta}_N \ e^{-\beta E_{\mathrm{Vega}}(\vec{\eta}_2,\ \cdots,\ \vec{\eta}_N)} \end{aligned} \tag{4.39}$$

It is easy to verify that the Jacobian of the change of the coordinates is one. It is also easy to see that the translationally invariant potential depends only on $\vec{\eta}_2, \cdots, \vec{\eta}_N$, but not on $\vec{q}_1$. In the paper by Vega and Noya, they adopt the thermodynamic integral method and computed Q_{Vega} by integrating from an Einstein molecule (instead of from an Einstein crystal as in the Frenkel-Ladd method).

Chapter 5: Introduction to Engineering Maxwell's Demons

5.1 The second law of thermodynamics and Maxwell's demon

The second law of thermodynamics can be equivalently stated in different ways for different physical processes. Clausius's statement of the second law of thermodynamics reads: "*Heat can never pass from a colder to a warmer body without some other change, connected therewith, occurring at the same time.*" [41]. Planck's statement reads: "*It is impossible to construct an engine which will work in a complete cycle, and produce no effect except the raising of a weight and cooling of a heat reservoir.*" [42] These statements are intuitively straightforward in the macroscopic world: in the winter, a hot cup of coffee left outside will finally become cold; in the summer, a can of Coke taken from a refrigerator will be eventually heated by the air; an air conditioner that transfers heat from the nice and cool office to the outside functions only when its compressor continuously performs work, consuming electrical energy from the power grid.

However, even at thermal equilibrium, atoms and molecules undergo erratic thermal fluctuations that are not negligible at the microscopic level. On rare occasions, such fluctuations can briefly transfer heat from a colder system to a hotter system. In this sense, microscopic thermal fluctuations appear to be able to violate

the second law of thermodynamics. This observation hints that one could possibly systematically violate the second law by manipulating or rectifying fluctuations at the microscopic level. In 1871, Maxwell [43] proposed a hypothetical, ethereal being that apparently violates the Clausius statement of the second law. In Maxwell's words "*... a being whose faculties are so sharpened that he can follow every molecule in its course, whose attributes are still as essentially finite as our own, would be able to do what is at present impossible to us.*" This being is now called Maxwell's demon.

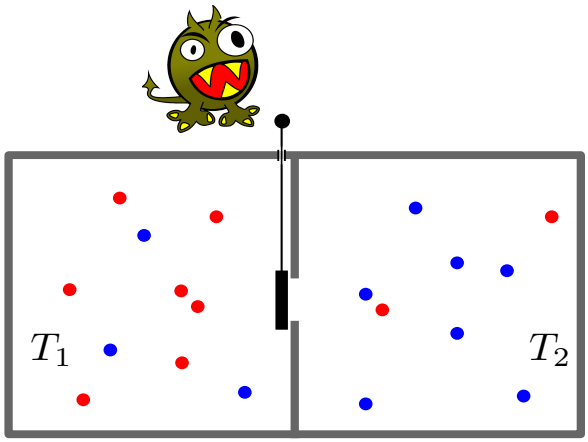


Figure 5.1: A sketch of Maxwell's demon. The red spheres represent particles with higher kinetic energy, and the blue ones with lower kinetic energy. Initially, the temperatures of the two containers are the same. Over time, the demon makes the left container hotter and the right cooler ($T_1 > T_2$). No work is done when the demon switches the door on the partition.

Maxwell's demon is briefly reviewed in Fig. 5.1. A container of gas is isolated from the environment. An adiabatic partition is inserted in the middle of the container. The partition has a tiny hole in it that allows gas particles to travel through. A door is attached to the hole, controlled by Maxwell's demon. The imagined being observes the motions of the gas particles and controls the door at the hole. If it observes a fast particle that travels to the hole from left to right, it shuts the door; if

it observes a slow particle traveling from right to left, it shuts the door. Over time, the left chamber accumulates particles that move faster than average particles and the right chamber accumulates slower particles. In other words, a directed heat flow is generated between two chambers, and over time one chamber becomes cooler and the other becomes hotter, without any other change.

If an imagined being could violate Clausius's statement of the second law, then in principle one can imagine another being that can violate Planck's statement of the second law. In other words, one could cook up an ethereal being that extracts useful mechanical work from a single heat bath without causing other changes. A good example is Szilard's Engine, which appears to be a perpetual motion machine of the second kind [44].

The Szilard Engine is sketched in Fig. 5.2. It consists of a single gas particle in a cylinder, a demon, a source of work, and a heat bath, and it works in cycles. The system is kept in contact with the heat bath. At the beginning of each cycle, a partition is inserted at the middle of the cylinder. Then, the demon measures the location of the gas particle (left or right). Based on the measurement outcome (left or right), an external force is attached against the partition (from the left or from the right). As the partition is released, the expansion of the gas pushes the partition and work against the external load. At the end of the cycle, the volume of the gas returns to its initial value. Over many cycles, if all of the measurement outcomes are truthful, then the system continuously works against the external load (by lifting a mass) while cooling a single heat bath. The maximum energy extracted from the heat bath per period is $k_B T \ln 2$, which is achieved when the expansion takes place

quasi-statically and isothermally. This engine violates Planck's statement of the second law.

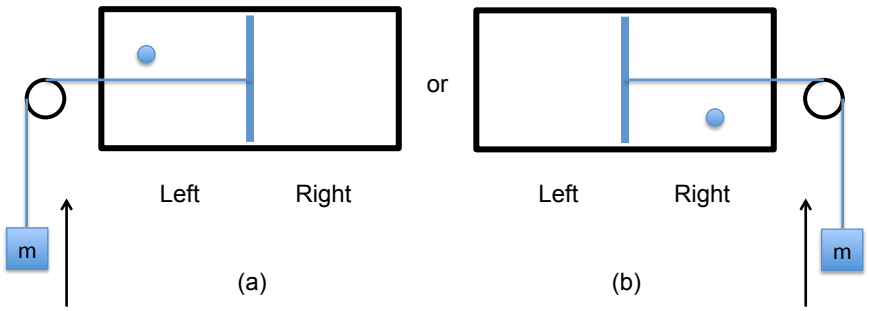


Figure 5.2: A sketch of Szilard's engine. Depending on the position of the particle (left or right), a string connected to a mass is attached to left or right side of the partition. As the gas expands from half the cylinder to the whole cylinder, the mass is lifted and positive work is done. The energy used to lift the mass against gravity comes from the surrounding heat bath (not shown).

5.2 Physical device as Maxwell's demon

Since both Maxwell's demon and Szilard's engine involve ethereal demons, it is not surprising that they can violate the second law of thermodynamics. It is natural to ask whether one can find a purely physical system that works like a Maxwell's demon.

The pursuit of an autonomous physical device that works as Maxwell's demon traces back to the early 1900s. In 1900 Lippmann [45] stated that if a ratchet and pawl device could be built at the nanoscale, then it could be used as a Maxwell's demon and violate the second law. In 1912, Smoluchowski [46] analytically showed that even if such a device existed, it could not violate the second law. Later in the 1960s, Feynman extended this discussion in his famous lecture [47]. Here let us briefly review the ratchet and pawl model (see Fig. 5.3). In this model, an axle

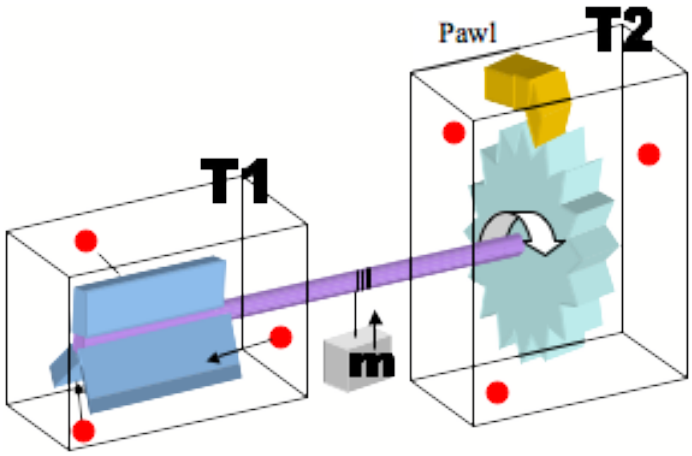


Figure 5.3: A Feynman-Smoluchowski ratchet. This figure is taken from [2].

connects a ratchet and pawl on one side to a set of paddles on the other side. The paddles are immersed in a container of gas particles at temperature T. The thermal fluctuations of the gas particles kick the paddles and generate a random rotation around the axle. At the same time, the ratchet and pawl on the other end of the axle only allows the axle to rotate in one direction. If this ratchet and pawl works faithfully, over time, the thermal fluctuations of the gas particles are rectified to a biased rotation that could be used to lift a small mass against gravity. As pointed out by Feynman, this device could only function in this manner when the ratchet and pawl have a different temperature from the temperature of the paddles (the gas). When the ratchet and pawl are maintained at the same temperature as the paddles, the thermal fluctuations of the pawl would result in constant failures and it could no longer rectify the brownian rotation of the paddles into a directed rotation. Similiar models have been studied analytically in reference [48] and verified experimentally in reference [49].

To summarize, although the ratchet and pawl is a purely physical and au-

tonomous device, it cannot violate Planck's statement of the second law. As shown by Feynman, in order for the device to deliver useful work, it must be equipped with two heat baths at different temperatures, and in this case the device reduces to a heat engine.

5.3 Landauer's principle and information engine

Can one design a hypothetical, autonomous physical device that works in cycles to extract energy from a single heat bath and performs work, by rectifying thermal fluctuations? If yes, how can such a device be reconciled with the second law?

To answer these questions, let us take Szilard's engine (at maximum efficiency) as an example. In each cycle, $k_B T \ln 2$ of heat is absorbed from the heat bath and thus its entropy decreases by $k_B \ln 2$. Note that when the system completes a cycle, it returns to its initial state without any change of entropy. Apparently the total entropy change of the system and the heat bath is $-k_B \ln 2$, which seems to contradict the second law. To reconcile with the second law, there must be an entropy increase of at least $k_B \ln 2$ per cycle somewhere else. In the following argument, originally due to Bennett [50,51], this entropy increase is associated with writing information to a physical memory register. Specifically, during each cycle, Szilard's engine measures the position of the particle (left/right), thereby gathering one binary digit (bit) of information. Bennett argued that if the device is a physical gadget rather than an ethereal demon, then this information must be written to a

physical memory register. Therefore, during each cycle, the device needs to write one bit of information to record the measurement outcome. Assume the information is written as a binary sequence of 0's and 1's on a tape that is initialized as a sequence of pure 0's. After n cycles, the Shannon entropy of the memory increases by $n \ln 2$, as the state space of the n bits tape expands from a unique state (purely 0's) to a set of 2^n equally possible states (all possible combinations of 0's and 1's). If we equate the Shannon entropy with thermodynamic entropy, we find that the entropy increase of the memory tape compensates the entropy decrease of the bath. With the entropic change of a memory register taken into consideration, Szilard's engine is thus reconciled with second law of thermodynamics.

Since the 1960s, Landauer [52], Penrose [53], and Bennett [50,51] have pointed out the connection between information processing and thermodynamics. Landauer's principle [52] states that it takes at least $k_B T \ln 2$ of energy dissipated into the heat bath to erase one bit of information from a memory register, where k_B is Boltzmann's constant, and T is the temperature of the bath. For a Szilard's engine that works in complete cycles, at the end of each cycle, both the system and the memory need to be restored to their initial states. In each cycle one bit of information is written on the memory. To restore this memory, at least $k_B T \ln 2$ of energy needs to be dissipated into the heat bath, according to Landauer's principle. In other words, although a Szilard's engine works in cycles and extract energy from a single heat bath to perform work, it cannot provide more work than the energy cost of restoring the memory register.

Landauer's principle provides the theoretical basis to reconcile Maxwell's de-

mon with the second law. Still, it is an interesting engineering task to actually design an autonomous model, which extracts work from a single heat bath, at the cost of writing information on a memory register. In 2012, Mandal and Jarzynski proposed such a model in the discrete stochastic scheme [3]. (See Fig. 5.4.) Their model consists of a demon that is represented by a three-state system, whose states are denoted by A, B, and C. The demon is provided with a memory tape consists of bits (0 and 1) traveling from left to right past the demon, and each bit interacts with the demon one after another. The stochastic transitions between states of the demon (A and B, and between B and C) are allowed in both directions at thermal fluctuation. However, a transition of the demon from state A to C must be accompanied by a bit switch from state 1 to state 0, and a transition from C to A must be accompanied by a bit switch from 0 to 1. Over time, if the demon is provided with a clean memory tape (e.g. ...000...), it generates directed rotation (counter-clockwise), which could be used to lift a mass against gravity. Other Maxwell's demon models have been designed and studied in [54–60].

In Chapter 6 and Chapter 7, we propose explicit continuous mechanical devices that work in a manner similar to the MJ model.

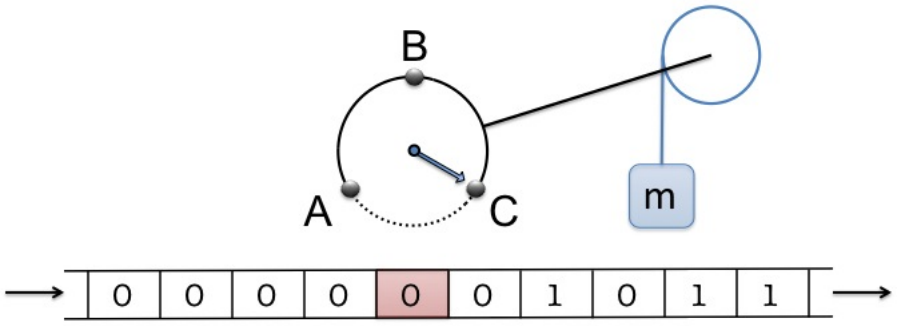


Figure 5.4: Mandal-Jarzynski model of Maxwell's demon. A tape of binary sequence travels from left to right past the demon. If the incoming bits are all 0's, the demon favors a counter clockwise rotation, which can be harnessed to lift a mass against gravity. This figure is taken from reference [3].

Chapter 6: Mechanical Maxwell's Demon

We aim to engineer an explicit mechanical device that works autonomously as a Maxwell's demon. The work described in this chapter is done in collaboration with Dr. Dibyendu Mandal, and was published as a Quick Study in Physics Today in 2014 [61].

6.1 Designing a mechanical Maxwell's demon

A snapshot of the device is shown in Fig. 6.1. In this section, we describe its components and their rules of motion. Refer to [62, 63] for videos of the device.

Similar to the MJ model, the whole system is immersed in an infinitely large heat bath, which is modeled as a dilute gas equilibrated at temperature T.

A stream of information bits is modeled by a concrete physical system – a sequence of paddle rotors along a central axle. There are two blocking bars (shown as vertical red lines) located at angles 0 and π, that separate the paddles into binary states. Each paddle rotates frictionlessly and is denoted by its angular position θ_B (see Fig. 6.1). When $0 < \theta_B < \pi$, the paddle is considered as a bit in state 0; when $\pi < \theta_B < 2\pi$, the paddle is considered as a bit in state 1. Each paddle undergoes random and erratic rotations due to the collisions with the gas particles from the

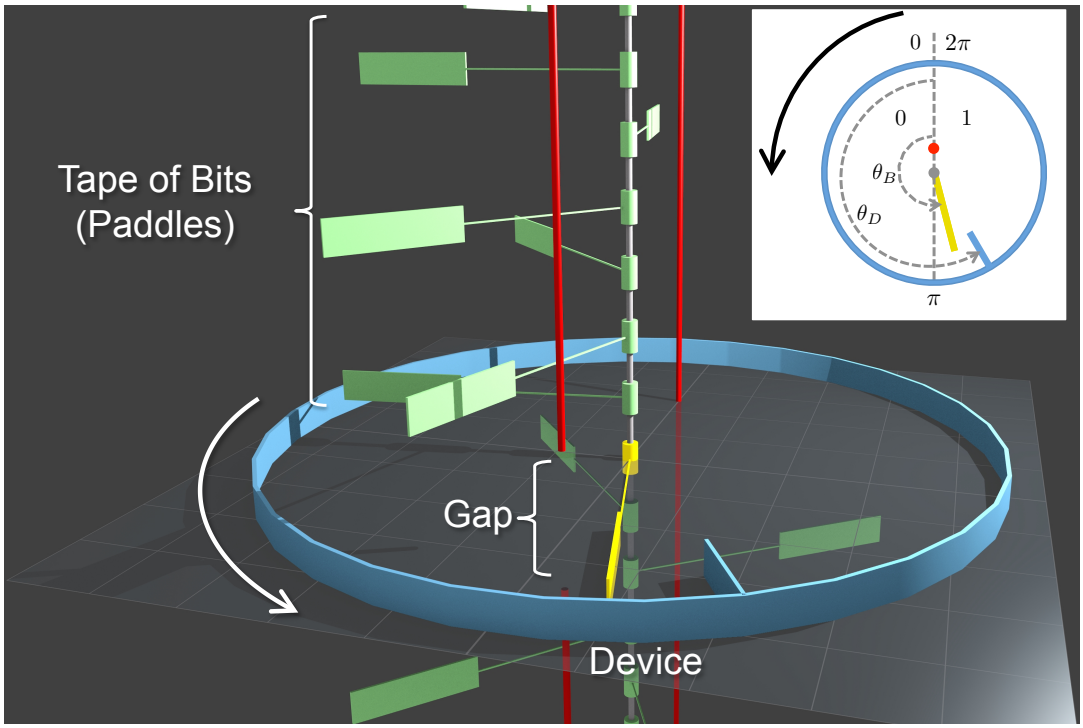


Figure 6.1: Snapshot of our mechanical Maxwell's demon. A sequence of green paddles are divided by two red blocking bars into binary states – left (0) and right (1). As the paddles (bits) collectively move down along the central axle, each bit passes the vertical range of the device (blue ring) for the same finite amount of time, during which it is able to switch states by rotating through the gap in the red blocking bar in the foreground. The interacting bit is highlighted as yellow in this snapshot, and it can collide with the blade attached to the device. We claim that if the incoming bits are in a pure sequence of 0's, then the device is biased to favor counterclockwise (CCW) motion, which can be harnessed to perform work against an external load. Due to symmetry, an incoming sequence of 1's would favor the device for clockwise (CW) rotation. Note that everything is immersed in a heat bath consisting of gas particles, and thus the paddles and the demon undergo random rotation about the central axis. A top view of the composite system (demon and interacting bit) is shown at the upper right corner.

heat bath; simultaneously, all of the paddles are collectively carried down by the axle (gray cylinder in Fig. 6.1) at a constant speed. Because of the blocking bars, the bits cannot switch state, unless they are at the vertical range of the gap located on the red blocking bar at $\theta = \pi$, when they can rotate through the gap. The collective downward motion of the paddles and the central axle is frictionless.

In place of the 3-state system that appears in the MJ model, our mechanical demon (the device) is a rigid blue ring with a blue blade attached on it (see FIG 6.1). The device is forbidden to translate or wobble, but it is allowed to rotate freely around the central axis. We denote the state of the device by the angular position of its blade, θ_D. Due to collisions with the gas particles, the device undergoes random rotation. The blade of the device can also collide with the paddle (bit) that is within the vertical range of the gap. The collision occurs when $\theta_D = \theta_B + 2N\pi$, for any integer N. We put the device at the vertical range of the gap and engineer the size of each part such that at any time, there is one and only one paddle that is within the vertical range of the gap, and at the same time, it is within the vertical collision range with the blade of the device. Such a bit is called the *interacting bit*. As the central axle moves down, each bit spends the same amount of time (which we denote τ^{bit}) as an interacting bit. Note that the blade of the device is too short to be blocked by the red blocking bar located at $\theta = 0$.

An external load can be attached to the device, such that work is extracted when the device systematically rotates in a particular direction (CW or CCW). We denote the load by torque Γ, and the sign of Γ is positive if the external load favors a CW rotation. In other words, if the device rotates in the CCW direction, then a

mass is lifted against gravity.

6.2 Modes of the device

Assume the bits are statistically independent of one another when they arrive. The entropy per bit of the incoming bit stream is

$$S_{\text{in}} = -k_B\big(P_{\text{in}}(0)\ln P_{\text{in}}(0) + P_{\text{in}}(1)\ln P_{\text{in}}(1)\big) \tag{6.1}$$

where $P_{\text{in}}(0)$ is the probability that a given incoming bit is in the 0 state, and $P_{\text{in}}(1)$ is defined similarly for the 1 state. The maximum entropy per bit is achieved when $P_{\text{in}}(0) = P_{\text{in}}(1) = 0.5$. The device is expected to function in 3 different modes depending on the following parameters – the distribution of 0's and 1's in the incoming sequence, the magnitude of the external load, and the temperature of the heat bath. In this section, we qualitatively describe these modes.

6.2.1 The engine mode

In the first mode, the device works as Maxwell's demon, converting heat into work while writing information onto a clean memory tape (or a low-entropy sequence). For simplicity, let us first assume that there is no external load. Also, assume that the incoming tape is a pure sequences of 0's (i.e....000...). In such a setup, we argue in the next two paragraphs that the demon preferentially undergoes a counter-clockwise (CCW) motion. Once systematic CCW rotation is achieved, one can attach a small load (positive Γ) to the demon. Here the load needs to be

small enough so that it does not stall or reverse the systematic CCW rotation. In this situation, the demon performs work as it lifts the load and cools the heat bath, while writing information on the clean tape.

In Fig. 6.2(a), we sketch the top view of the demon and the interacting bit. Since we assume the incoming bits are all 0's, every incoming bit arriving at the demon begins with an angle that is taken randomly from 0 to π (the left half circle). When a paddle arrives at the position of the demon (or becomes a interacting bit), it is able to switch its state by passing through the gap, shown in Fig. 6.2 as a dashed red circle located at $\theta = \pi$. From an ensemble point of view, these incoming paddles confined to the left half circle (state 0) tend to expand spontaneously from the initial half circle to a full circle, which can only occur one at a time when a bit becomes the interacting bit. By design, the expansion of the paddles is CCW. As a result the demon, on average, is pushed by the expanding paddles in a CCW rotation.

Similarly, in Fig. 6.2(b), one can be convinced that if the incoming sequence of bits contains only 1's, the isothermal expansion of the bit paddles favors CW rotation. As a result, the demon favors CW rotation when it is provided with a pure tape of 1's. Notice that in this case, to have the demon perform positive work against the load, one needs to reverse the direction of the external load so that $\Gamma < 0$.

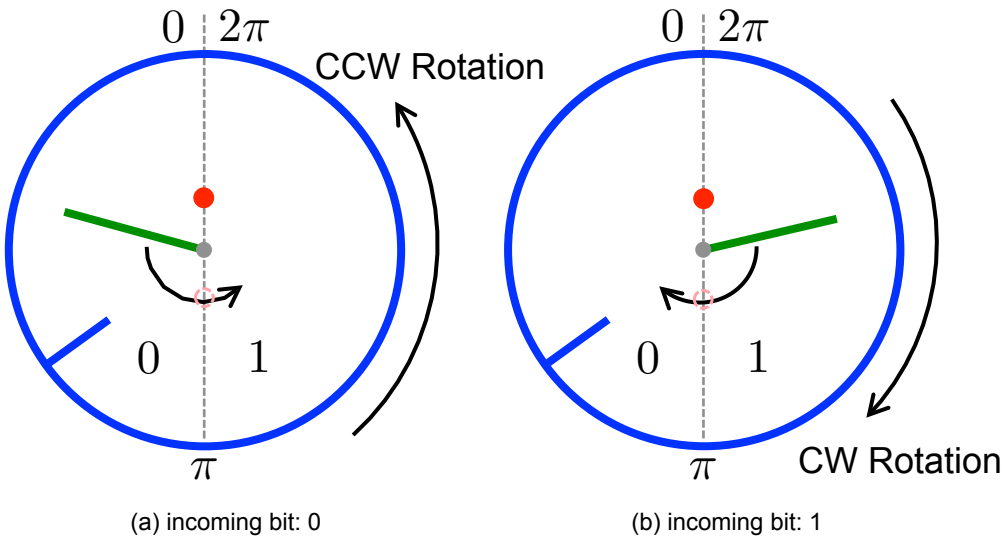


Figure 6.2: Engine Mode

6.2.2 The eraser mode

As discussed in Sec. 5.3, Landauer's principle states that erasing one bit of information from a memory register costs at least $k_B T \ln 2$ of energy to be dissipated into the heat bath. Similar to the MJ model, our device can work as a Landauer's eraser. Specifically, if the incoming sequence of bits features a random combination of 0's and 1's, by dissipating work into the heat bath our device can erase the randomness and produce a low-entropy outgoing sequence (e.g. purely 0's or purely 1's). To illustrate this mode qualitatively, let us assume that the incoming sequence is a random mixture of 0's and 1's, and the external load is very heavy. We can show that depending on the direction of the heavy load ($\Gamma \gg 1$ or $\Gamma \ll -1$), the outgoing sequence is either a pure sequence of 0's or a pure sequence of 1's. In other words, the entropy of the sequence decreases as work done by the external load is dissipated into the heat bath.

Fig. 6.3 (a) and (b) shows the top view of the demon and the interacting bit when the heavy load is applied from different directions. Note that the rotation of the demon is not directly interfered by the red blocking bar (at $\theta = 0$), but indirectly blocked by the paddle that is stopped by the blocking bar. If the heavy load favors a CW rotation ($\Gamma > 0$), the paddle and the demon's blade are stalled on the left side of the blocking bar, and the paddle is pinched at state 0 until it leaves the demon (Fig. 6.3(a)). As a result, the outgoing sequence is purely 0. If the heavy load favors a CCW rotation ($\Gamma < 0$), the paddle and the demon's blade is stalled on the right side of the blocking bar, and the paddle is pinched at state 1 until it leaves the demon (Fig. 6.3(b)). As a result, the outgoing sequence is purely 1. In both cases, the external load is lowered, and the gravitational energy of the load is dissipated into the surrounding gas.

As a result, the energy from the load dissipates into the thermal bath, and the random incoming sequence is erased and cleaned into a pure sequence of 0's or 1's, depending on the direction of the load applied.

6.2.3 The dud mode

There are situations where the demon is not able to work as an engine nor an eraser. This mode is called the dud mode. In such a mode, work done by the load is dissipated into the bath, at the same time, entropy of the sequence increases as it passes the demon. When the external load is too heavy, a demon originally at the engine mode could fall in the dud mode. Similarly, a demon in the eraser mode

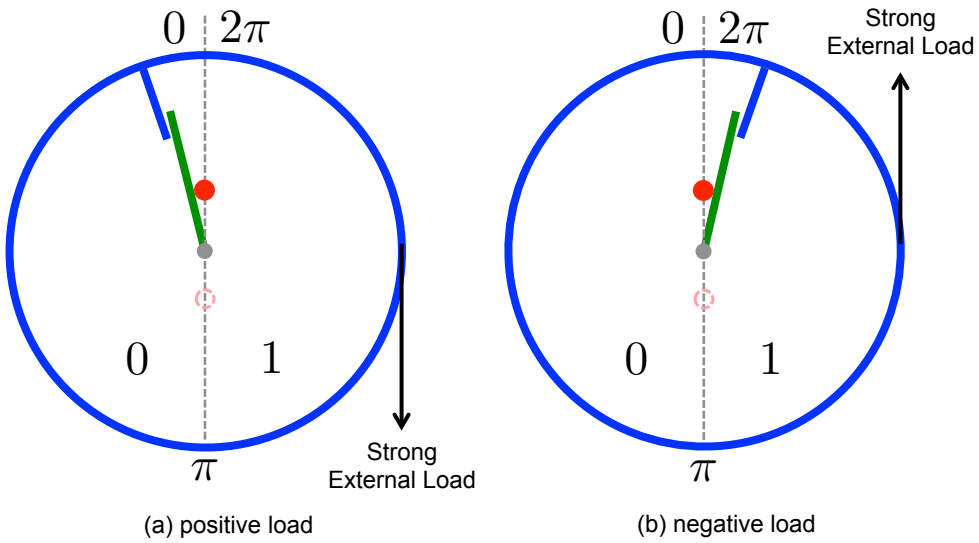


Figure 6.3: Eraser Mode

could also fall in the dud mode if the strength of the load is too weak.

6.3 Quantitative analysis of the mechanical Maxwell's demon

In this work, for simplicity, we prepare the incoming sequence of bits such that all of the incoming bits are statistically independent from each other, and the incoming sequence is characterized by the excess ratio of 0's defined by

$$\delta \equiv P_{\text{in}}(0) - P_{\text{in}}(1) \tag{6.2}$$

As the axle carries the sequence downward past the demon, each incoming bit spends the same amount of time at the vertical level of the demon, namely the vertical range of the gap on the blocking bar. We name this time interval the *interaction interval*, and denote it by τ^{bit}. During each interaction interval, the interacting bit is able to collide with the demon, and is able to switch its state between 0 and 1. We denote

the ratios of 0's and 1's of the outgoing tape by $P_{\rm out}(0)$ and $P_{\rm out}(1)$.

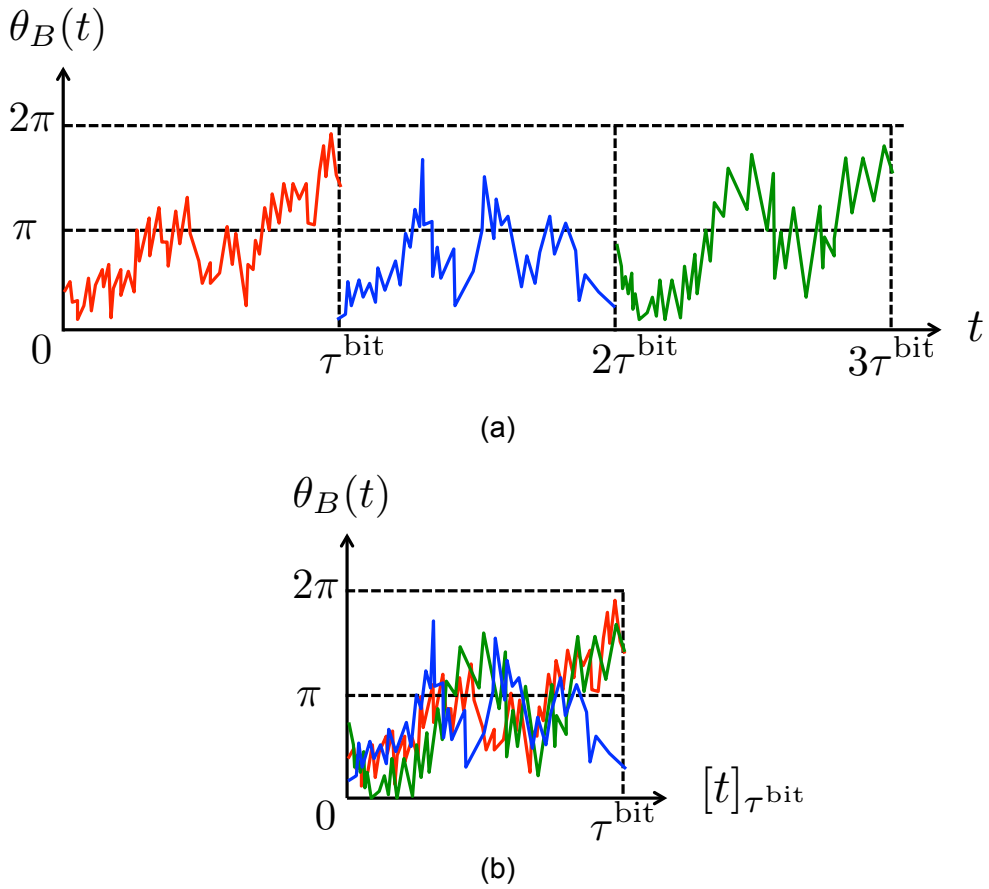


Figure 6.4: Sketched in (a) is the trajectory of the interacting bit during 3 interaction intervals. In each interaction interval, the interacting bit performs random rotation, and is allowed to explore the whole range from 0 to 2π. At the end of each interval, a new bit (in state 0) comes to replace the old one, with a initial value confined in $(0, \pi)$. In (b), we reset the clock at the end of every interaction interval and plot the trajectories of $\theta_B(t)$. The trajectories shown in (b) illustrates the expansion from $(0, \pi)$ to $(0, 2\pi)$ during each interval. In other words, after interacting with the demon, each incoming bit initially in state 0 can be found in either state 0 or state 1.

In this section, we describe the analytical method for solving the average work performed against the load per bit, and the Shannon entropy change of the sequence per bit. Let us start by separately considering the dynamics of the demon and the interacting bit. During each interaction interval, of duration $\tau^{\rm bit}$, the demon (θ_D)

evolves under random motion due to the thermal fluctuations (collisions with gas particles) and the interaction with the interacting bit (collisions with the interacting paddle). During the interaction interval, the interacting bit $\theta_B(t)$ also evolves randomly due to the collisions with the gas particles and the demon. At the end of the interaction interval, a new bit replaces the old one, which we call *bit renewal*. At bit renewal, the value of $\theta_B(t)$ changes discontinuously– the new bit arrives in a state that is independent of the current bit. After bit renewal, the state of the outgoing bit is not able to change, and is determined by its angular position right before it leaves the demon. There is no physical discontinuity in the motion of the paddles (θ_B) but the angle of the interacting bit is instantly updated to a new value at the beginning of each interval. We sketch the trajectory of $\theta_B(t)$ in Fig. 6.4.

In the following, let us focus on the dynamics of the *composite system* of the demon plus the interacting bit in the 2-dimensional configuration space. (See Fig. 6.5.) Due to the hard collisions between the paddle (bit) and the red blocking bar (at $\theta = 0$ or equivalently, 2π), the angle of the bit cannot pass 0 or 2π. Additionally due to the hard collisions between the paddle (bit) and the blade on the ring (demon), the blade of the demon and the paddle cannot cross each other. These restrictions are represented by hard boundaries in the configuration space of the composite system. Shown in Fig. 6.5, these hard boundaries (at $\theta_B = 0$ and $\theta_B = \theta_D + 2N\pi$) form a "ladder" – a set of adjacent parallelograms that form a 1-dimensional lattice of cells. Within each interaction interval, the composite system starting within one cell cannot cross the boundaries of that cell, i.e. the random motion of the composite system is strictly confined within the cell. At the instants

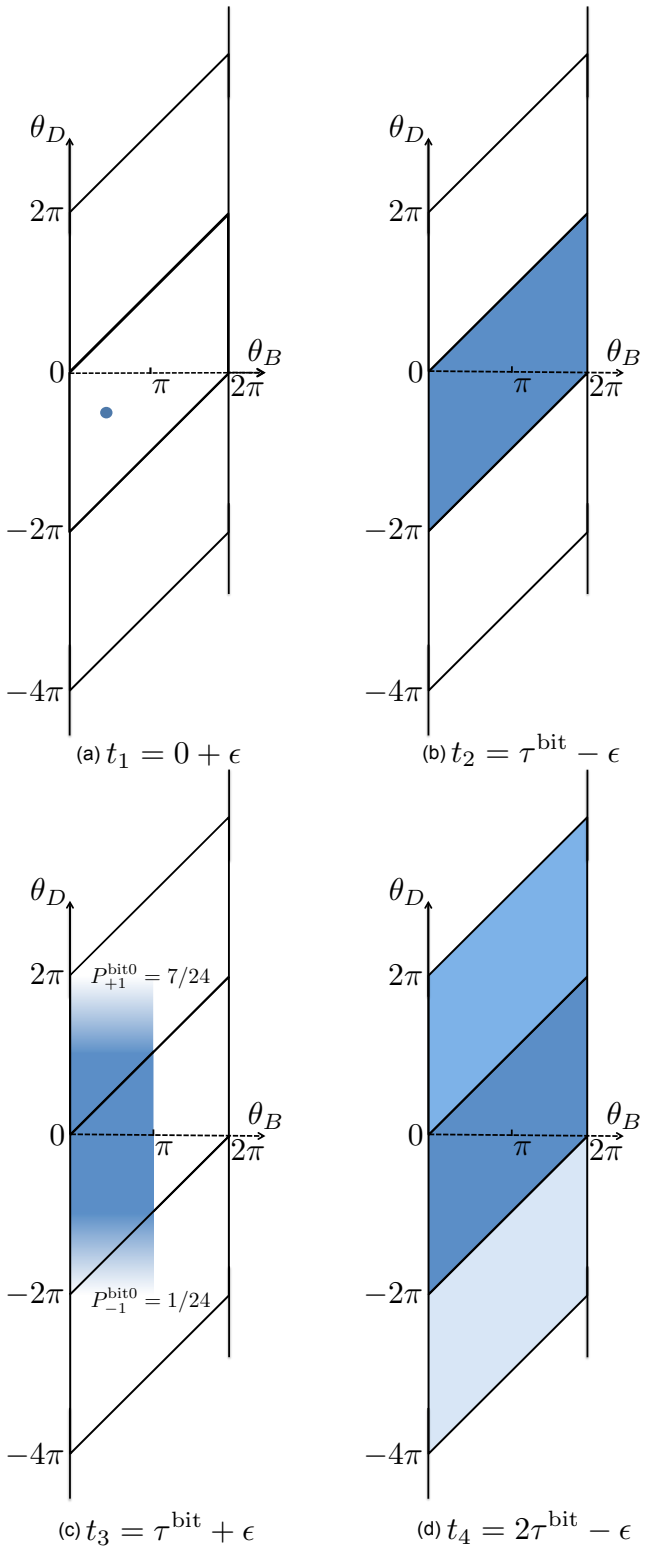


Figure 6.5: Configuration space of the composite system (θ_B, θ_D). For simplicity, we assume no external load is attached and $\tau^{\text{bit}} \gg \tau^{\text{eq}}$. The solid black lines represent hard wall boundary conditions. The four density plots sketch $p(\theta_B, \theta_D, t)$ – the probability distributions of the composite state at different times. (ϵ is a positive infinitesimal number.)

of bit renewals, however, the composite system is able to jump from one cell to its adjacent cells. These instantaneous jumps correspond to the discontinuous updates of $\theta_B(t)$ that occur when an old bit is replaced by a new interacting bit. The value of θ_D does not change during bit renewal.

Immediately, one can map the dynamics of the composite system to a random walk over a one-dimensional lattice (along the θ_D axis), where the system can jump from its current cell to adjacent cells at the end of each interaction interval. Note that the probability of jumping up or jumping down is conditioned on the state of the new incoming bit, and depends on the probability distribution of the composite system right before the bit renewal. Assuming that a new bit arrives in state 0, let us denote the probability of jumping up by one cell as P_{+1}^{bit0}; the probability of jumping down by one cell as P_{-1}^{bit0}, and the probability of staying at the original cell as $P_{\text{stay}}^{\text{bit0}}$. On average, the composite system jumping up by one cell corresponds to a full-circle CCW rotation of the demon, and jumping down by one cell corresponds to a full-circle CW rotation of the demon. Additionally, by integrating over the θ_D degrees of freedom, one can compute the probability of the states of the outgoing bits from the distributions of the composite system right before bit renewal.

Let us denote the relaxation time of the composite system as τ^{eq}, and assume $\tau^{\text{bit}} \gg \tau^{\text{eq}}$. Under this assumption, the distribution of the composite system relaxes to a Boltzmann's distribution before each bit renewal, allowing us to analytically solve for the performance of the demon.

First let us consider a simple case as an example, where the magnitude of the external load is zero, and the incoming sequence is purely 0. We sketch the

evolution of an ensemble of composite systems in Fig. 6.5. Initially at $t_1 = 0+\epsilon$, the composite state (shown as a blue dot) is prepared within one cell (see Fig. 6.5(a)). As the ensemble of composite systems evolve, at time t_2, they relax into a Boltzmann distribution (or a uniform distribution in the absence of external load) restricted to the cell (see Fig. 6.5(b)). At the end of the interaction interval, a new bit in state 0 comes to replace the old interacting bit. This renewal at t_3 instantly maps θ_B to a random value taken from $0 < \theta_B < \pi$, with θ_D held fixed. As a result, a portion (7/24) of the composite states jump up by one cell, and 1/24 of the composite states jump down by one cell (see Fig. 6.5(c)). After another interaction interval of relaxation, at t_4, the ensemble of composite systems relax into an equilibrium distribution within 3 adjacent cells (see Fig. 6.5(d)). Clearly, with this set of parameters, the demon favors CCW rotation, with an average rotation angle $2\pi(7-1)/24 = \pi/2$ per period $\tau^{\rm bit}$. Additionally, by integrating over θ_D the probability distribution of the composite state at time t_2, one can find the probabilities of states for the bit that left the demon. In this case the outgoing bit is equally likely to be in either state 0 or 1.

For a more general setup, i.e. with nonzero load and arbitrary probabilities of 0's and 1's in the incoming sequence, we can analytically solve for the performance of the demon as long as $\tau^{\rm bit} \gg \tau^{\rm eq}$. First of all, let us write down the general form of probability distribution of the composite system right before the bit renewal (at t_2 in Fig. 6.5):

$$p(\theta_B, \theta_D, t_2) = C \cdot \Theta_{t_2}(\theta_B, \theta_D) e^{-\beta\Gamma\theta_D} \tag{6.3}$$

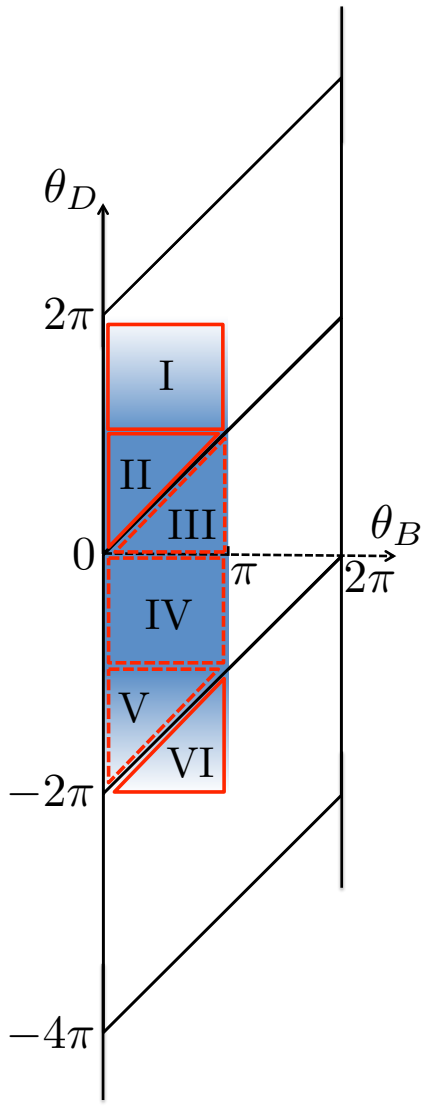


Figure 6.6: The probability distribution of the composite system right after bit renewal with new bit at state 0. We divide the support of the distribution into 6 pieces for the sake of argument.

where C is a normalization constant and $\Theta_{t_2}(\theta_B, \theta_D)$ is an indicator function. Here $\Theta_{t_2}(\theta_B, \theta_D) = 1$ when (θ_B, θ_D) belongs to the highlighted parallelogram cell in Fig. 6.5(b), and it equals to 0 otherwise. Then, immediately after the bit renewal (assuming the new bit arrives in state 0), we can write the probability distribution (see Fig. 6.5(c) and Fig. 6.6)

$$p(\theta_B, \theta_D, t_3) = \frac{\beta^2\Gamma^2}{4\pi \sinh^2(\pi\beta\Gamma)} \cdot \Theta_{t_3}(\theta_B, \theta_D) \cdot (2\pi - |\theta_D|) \cdot e^{-\beta\Gamma\theta_D} \tag{6.4}$$

where $\Theta_{t_3}(\theta_B, \theta_D)$ is a indicator function corresponding to the shaded region in Fig. 6.6(c). Precisely, $\Theta_{t_3}(\theta_B, \theta_D)$ equals to 1 when $\theta_B \in (0, \pi)$ and $\theta_D \in (-2\pi, 2\pi)$, and 0 otherwise. As is shown in Fig. 6.6, we can integrate $p(\theta_B, \theta_D, t_3)$ piece by piece to find the probability of the system jumping up, down and staying at the same cell:

$$P_{+1}^{\text{bit0}} = \int_{\text{I\&II}} \mathrm{d}\theta_B \mathrm{d}\theta_D \; p(\theta_B, \theta_D, t_3) \tag{6.5}$$

$$P_{-1}^{\text{bit0}} = \int_{\text{VI}} \mathrm{d}\theta_B \mathrm{d}\theta_D \; p(\theta_B, \theta_D, t_3) \tag{6.6}$$

and

$$P_{\text{stay}}^{\text{bit0}} = \int_{\text{III\&IV\&V}} \mathrm{d}\theta_B \mathrm{d}\theta_D \; p(\theta_B, \theta_D, t_3) \; . \tag{6.7}$$

Using this strategy one can also calculate P_{+1}^{bit1}, $P_{\text{stay}}^{\text{bit1}}$, and P_{-1}^{bit1}. We summarize the analytical results below.

When a new bit arrives in state 0, the probabilities of the composite system jumping up by one cell (or completing a full CCW rotation), jumping down by one

cell, or staying at the original cell are

$$P_{+1}^{\text{bit0}} = \frac{\pi\beta\Gamma + e^{\pi\beta\Gamma}\left[2e^{\pi\beta\Gamma}(\pi\beta\Gamma - 1) - \pi\beta\Gamma + 2\right]}{\pi\beta\Gamma\left(e^{2\pi\beta\Gamma} - 1\right)^2} \tag{6.8}$$

$$P_{-1}^{\text{bit0}} = \frac{e^{3\pi\beta\Gamma}\left[e^{\pi\beta\Gamma}(\pi\beta\Gamma - 2) + \pi\beta\Gamma + 2\right]}{\pi\beta\Gamma\left(e^{2\pi\beta\Gamma} - 1\right)^2} \tag{6.9}$$

and

$$P_{\text{stay}}^{\text{bit0}} = \frac{e^{\pi\beta\Gamma}\left(\pi\beta\Gamma - 2\right) + e^{2\pi\beta\Gamma}\left[-4\pi\beta\Gamma + 2e^{2\pi\beta\Gamma} - e^{\pi\beta\Gamma}(\pi\beta\Gamma + 2) + 2\right]}{\pi\beta\Gamma\left(e^{2\pi\beta\Gamma} - 1\right)^2} \tag{6.10}$$

Alternatively, when the new incoming bit arrives in state 1, we find

$$P_{+1}^{\text{bit1}} = \frac{\pi\beta\Gamma + e^{\pi\beta\Gamma}(\pi\beta\Gamma - 2) + 2}{\pi\beta\Gamma\left(e^{2\pi\beta\Gamma} - 1\right)^2} \tag{6.11}$$

$$P_{-1}^{\text{bit1}} = \frac{e^{2\pi\beta\Gamma}\left(2\pi\beta\Gamma + 2\right) + e^{3\pi\beta\Gamma}\left[\pi\beta\Gamma\left(e^{\pi\beta\Gamma} - 1\right) - 2\right]}{\pi\beta\Gamma\left(e^{2\pi\beta\Gamma} - 1\right)^2} \tag{6.12}$$

and

$$P_{\text{stay}}^{\text{bit1}} = \frac{2e^{2\pi\beta\Gamma}\left[-2\pi\beta\Gamma + \pi\beta\Gamma\sinh(\pi\beta\Gamma) + 2\cosh(\pi\beta\Gamma) - 1\right] - 2}{\pi\beta\Gamma\left(e^{2\pi\beta\Gamma} - 1\right)^2} \tag{6.13}$$

In general, the incoming tape may contain a mixture of 0's and 1's, and we have assumed that each bit follows an independent and identical probability distribution. Observing the demon for a long time (over many periods), the evolution of the

composite system can be modeled as a discrete time random walk along θ_D, with discrete time step τ^{bit}. During each step τ^{bit}, the composite system has a probability to jump up by one cell (θ_D gaining 2π),

$$R_{\text{up}} = P_{\text{in}}(0) \cdot P_{+1}^{\text{bit0}} + P_{\text{in}}(1) \cdot P_{+1}^{\text{bit1}} \tag{6.14}$$

or jump down by one cell (θ_D losing 2π),

$$R_{\text{down}} = P_{\text{in}}(0) \cdot P_{-1}^{\text{bit0}} + P_{\text{in}}(1) \cdot P_{-1}^{\text{bit1}} \tag{6.15}$$

The random walk, if biased, would generate a net CCW or CW rotation of the device, which can be harnessed to perform work against external load. The average work done against the external load during each period can be found by simply multiplying the external torque to the average rotation of the demon per period:

$$\begin{aligned} W &= (2\pi\Gamma) \cdot (R_{\text{up}} - R_{\text{down}}) \\ &= \frac{2\delta \tanh\left(\frac{\pi\beta\Gamma}{2}\right) + 2}{\beta} - \\ &\frac{1}{2}\pi\Gamma \left[\operatorname{sech}^2\left(\frac{\pi\beta\Gamma}{2}\right) \left(\sinh(\pi\beta\Gamma) + \delta \right) + 2\coth\left(\frac{\pi\beta\Gamma}{2}\right) \right]. \end{aligned} \tag{6.16}$$

where δ denotes the randomness of the incoming sequence (see Eq. 6.2).

As a consistency check let us evaluate Eq. 6.16 in the following limits. When the external load is extremely small ($\Gamma \to 0$), then to leading order in Γ Eq. 6.16 reduces to $W = \delta\pi\Gamma/2$. As expected, no work is done in the absence of an external load. When the load is very large ($|\Gamma| \gg k_B T$), then during each interaction interval

we expect the demon to complete either a full circle of CW rotation (if $\Gamma > 0$) or a full circle of CCW rotation (if $\Gamma < 0$). In agreement with this expectation, Eq. 6.16 gives us $W = \mp 2\pi\Gamma$ in the limit of $\Gamma \to \pm\infty$.

Additionally, we can compute the ratio of 0's and 1's in the outgoing tape

$$P_{\text{out}}(0) = 1 - \frac{1}{e^{\pi\beta\Gamma} + 1} \quad , \tag{6.17}$$

$$P_{\text{out}}(1) = \frac{1}{e^{\pi\beta\Gamma} + 1} \quad . \tag{6.18}$$

Note that when there is no external load ($\Gamma = 0$), then the outgoing bits are in a statistical mixture with $P_{\text{out}}(0) = P_{\text{out}}(1) = 1/2$. When the load is very large, $|\Gamma| \gg k_B T$, then the outgoing bits are either all 0's (if $\Gamma > 0$) or all 1's (if $\Gamma < 0$), in agreement with the qualitative discussion in Sec. 6.2.2. Thus, we can compute the change of Shannon entropy of the sequence per period.

$$\begin{aligned} \Delta S_{\text{bit}} = &- \frac{\left(e^{\pi\beta\Gamma} + 1\right)\left[(\delta - 1)\log(1 - \delta) - (\delta + 1)\log(\delta + 1) + \log(4)\right]}{2\left(e^{\pi\beta\Gamma} + 1\right)} \\ &- \frac{e^{\pi\beta\Gamma}\log\left(1 - \frac{1}{e^{\pi\beta\Gamma}+1}\right) - \log\left(e^{\pi\beta\Gamma} + 1\right)}{\left(e^{\pi\beta\Gamma} + 1\right)} \end{aligned} \tag{6.19}$$

With these analytical solutions under the assumption that $\tau^{\text{bit}} \gg \tau^{\text{eq}}$, we can further obtain the phase diagram for the demon. By choosing different combinations of the parameters – temperature (T), external load (torque Γ), and the quality of the incoming tape (δ), we can find the device in engine mode, eraser mode or dud mode. In the engine mode, positive work is done against the external load. In the

eraser mode, the entropy change of the tape is negative. In the dud mode, negative work is extracted and the entropy of the memory tape increases. Note that the second law of thermodynamics does not allow the device to simultaneously function at the engine mode and the eraser mode. The phase diagram is shown in Fig. 6.7.

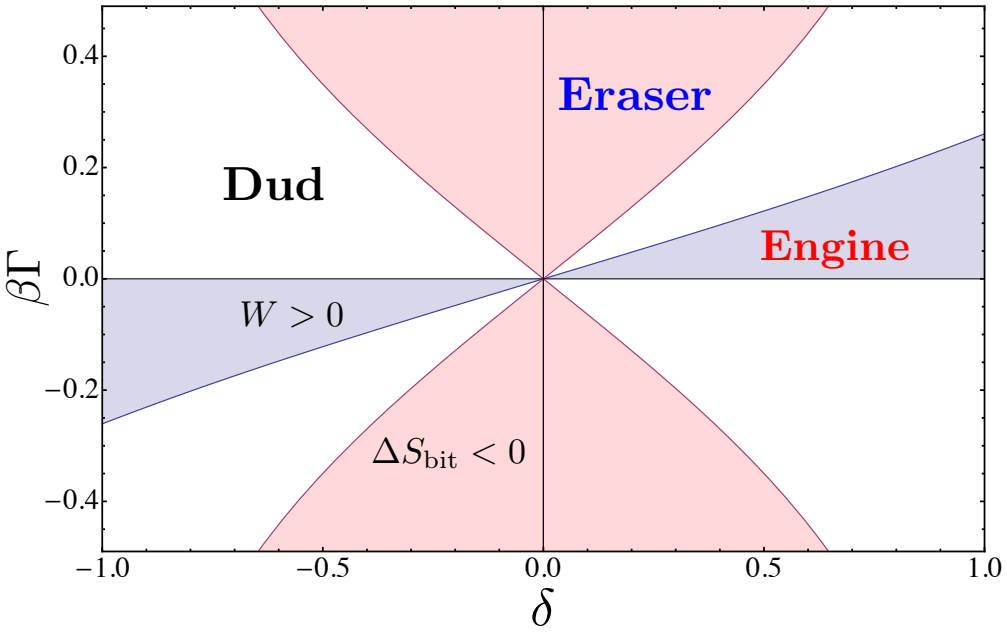


Figure 6.7: The phase diagram of the Maxwell's demon, assuming $\tau^{\rm bit} \gg \tau^{\rm eq}$.

6.4 Discussion

The model described in this chapter is an explicit autonomous mechanical device that extracts energy from a heat bath while writing information to a memory register. Without considering the change of information or the change of the Shannon entropy of the memory register, the device (in the engine mode) violates the second law of thermodynamics. Specifically, in such a case, energy is absorbed from the heat bath, and it is fully converted into mechanical work ($W > 0$). The only

change of entropy is

$$\Delta S_{\text{bath}} = \frac{Q_{\text{bath}}}{T} = \frac{-W}{T} < 0 \tag{6.20}$$

which contradicts the second law of thermodynamics. However, if we include the change of the information reservoir, or the memory sequence, the increase of the Shannon entropy of the sequence can be used to compensate the decrease of the entropy of the heat bath

$$\Delta S_{\text{bit}} > 0 \quad . \tag{6.21}$$

We can write the total entropy change per period as

$$\Delta S_{\text{total}} = k_B \Delta S_{\text{bit}} + \Delta S_{\text{bath}} = k_B \Delta S_{\text{bit}} + \frac{Q_{\text{bath}}}{T} = k_B \Delta S_{\text{bit}} - \frac{W}{T} \geq 0 \tag{6.22}$$

where the inequality follows from the second law. This immediately provides an upper bound on the amount of work one can extract from the heat bath

$$W \leq k_B T \Delta S_{\text{bit}} \quad . \tag{6.23}$$

Landauer's principle asserts that to erase one bit of information, one needs to dissipate at least $k_B T \ln 2$ of energy, on average. Our model can also work as an Landauer's eraser, where work is done while the entropy of the information sequence decreases.

In the following, we use efficiency to characterize the energy flow at different

modes of the demon. (See Fig. 6.8.) When the device functions as an eraser,

$$W < k_B T \Delta S_{\text{bit}} < 0 \tag{6.24}$$

and the efficiency is defined as

$$\eta_{\text{eraser}} = \frac{k_B T \Delta S_{\text{bit}}}{W} \leq 1 \tag{6.25}$$

When the device functions as an engine,

$$k_B T \Delta S_{\text{bit}} > W > 0 \tag{6.26}$$

and the efficiency is defined as

$$\eta_{\text{engine}} = \frac{W}{k_B T \Delta S_{\text{bit}}} \leq 1 \tag{6.27}$$

When the device functions in the dud mode,

$$W < 0 < k_B T \Delta S_{\text{bit}}. \tag{6.28}$$

In all modes, as long as the entropy change of the information sequence is included into consideration, the second law of thermodynamics is satisfied.

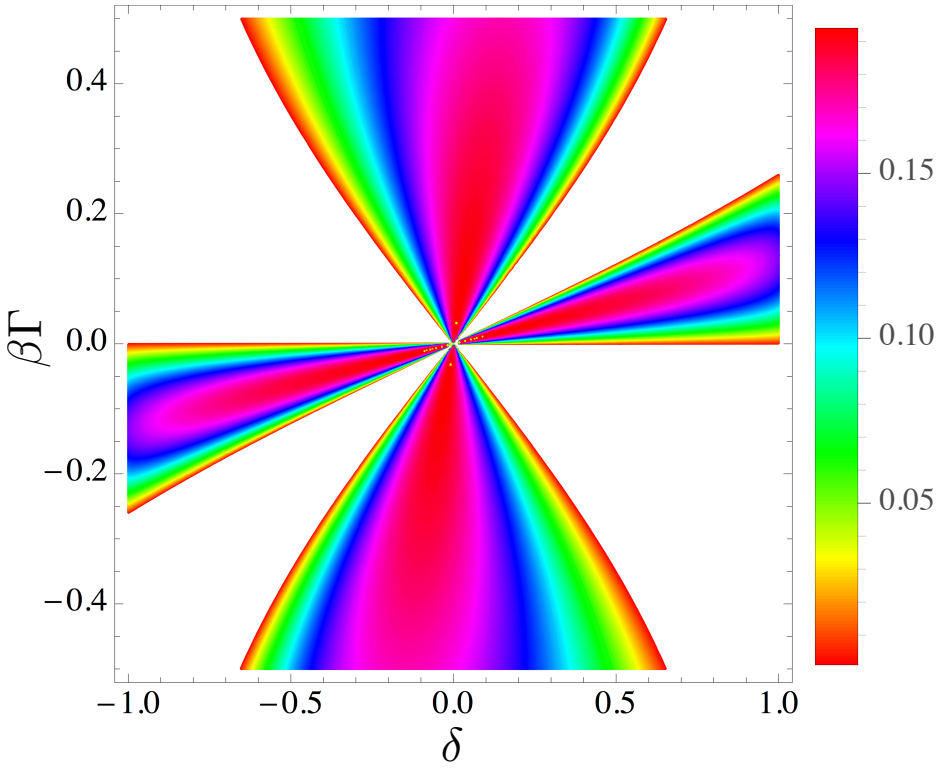


Figure 6.8: Efficiency plot of the demon, assuming $k_B = 1$, $\tau^{\text{bit}} \gg \tau^{\text{eq}}$.

Chapter 7: Programmable Mechanical Maxwell's Demon

7.1 Motivation

In chapter 6, we described an autonomous mechanical device that operates as an information engine. In that model, the ability to extract work from a bath directly relates to the fractions of 0's and 1's (or "cleanliness") in the incoming memory tape. We define a clean memory tape to be the one that allows the demon to extract the maximum amount of work from the heat bath to lift an external load. When $\Gamma > 0$, a sequence of all 0's is considered to be a clean memory. When $\Gamma < 0$, the maximum productivity is achieved when it is provided with the sequence ...1111..., hence a sequence of all 1's is considered to be clean.

Here we argue that, in addition to the two pure sequences (all 0's and all 1's), any sequence (even an irregular mixture of 0's and 1's) can be used as a clean memory, as long as each bit of the sequence is known ahead of time. From an entropic point of view, any determined sequence of 0's and 1's can be reversibly transformed to 0's and and they both have zero entropy (or zero information). Taking the binary representation of the number π as an example, although the sequence appears to be an irregular combination of 0's and 1's, it is still a unique well defined sequence without any uncertainty.

In principle, since we have constructed a Maxwell's demon that functions by writing information onto one specific clean sequence (all 0's), one should be able to engineer another device that functions similarly by writing onto any other clean sequence (e.g. the binary digits of π). The aim of this chapter is to find a mechanical Maxwell's demon that extracts work from a single heat bath by writing information to any pre-determined clean sequence (e.g. the binary representation of π, of $\sqrt{2}$, or anything else). To achieve this goal, the device must be pre-programmed with a reference sequence that defines its clean sequence.

7.2 Designing the programmable Maxwell's demon

In contrast with the demon engineered in chapter 6, which is designed only for one specific representation of a clean tape (...000...), this programmable demon is equipped with a programmable part that defines a reference sequence, for instance the binary representation of π. If the sequence of incoming bits matches the pre-programmed reference sequence, the device is expected to favor CCW rotation that can be used to perform work against an external load.

A snapshot of the device is shown in Fig. 7.1. Refer to [64] for a video of the device. The components and their interactions are described in the following paragraphs. The entire system is immersed in a heat bath, composed of gas particles equilibrated at temperature T. These gas particles collide elastically with all components of the system.

The sequence of information bits is modeled by a series of paddle rotors around

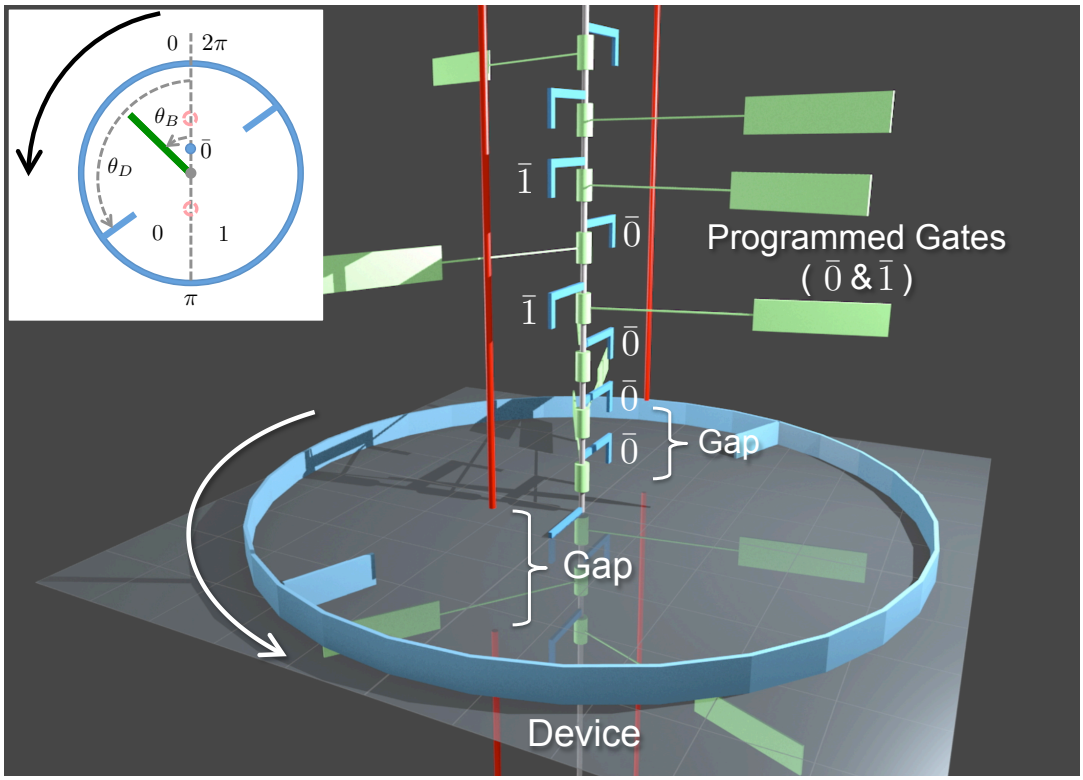


Figure 7.1: Snapshot of a programmable demon. A series of green paddles collectively move down along the central axle. The paddles are separated by the red bars into binary states, left (0) and right (1). Each bit falls in the range of the device (blue ring) for the same finite amount of time, when it is allowed to switch states. We claim that if the incoming bits $(000101\cdots)$ are in agreement with the programmed gates $(\bar{0}\bar{0}\bar{0}\bar{1}\bar{0}\bar{1}\cdots)$, then the device favors CCW motion, which can be used to attach an external load. A top view of the system is shown at the upper left corner.

the central axle. There are two red blocking bars (shown as vertical red bars) located at angle 0 and π, that separate the paddles into binary states. A paddle in the left half circle $\theta_B \in (0, \pi)$ represents a bit in state 0, and a paddle in the right $\theta_B \in (\pi, 2\pi)$ represents a bit in state 1. There are two types of motions for each paddle: each paddle undergoes random rotation due to the collision with the gas particles from the heat bath; additionally, all of the paddles are collectively carried down by the central axle at a constant speed. The bits are able to switch state, only when they arrive at the vertical range of the gaps located on the red bar at $\theta = 0$ and $\theta = \pi$. The collective downward motion of the paddles and the central axle is frictionless.

The programmable part of the system consists of a sequence of reference binary gates modeled by a set of rigid gates on the central axle that accompany the sequence of bits (paddles). These binary gates are the L-shaped blue bars extending perpendicularly from the central axle. The orientation of a gate is fixed at either $\theta = 0$ (in state $\bar{0}$) or $\theta = \pi$ (in state $\bar{1}$). These gates are rigid and are not allowed to rotate or wobble. The gates are evenly spaced and each gate accompanies one bit (paddle). As the central axle and the paddles move down, the gates are simultaneously carried down so they keep pace with the memory sequence. When a bit and its corresponding gate arrive at the vertical location of the gaps on the red bars, the paddle is able to switch its state only by passing through the gap that is not occupied by the gate. For example, if the gate is in state $\bar{0}$, the gap at $\theta = 0$ is blocked, and the bit can only switch its state by passing through the gap at $\theta = \pi$.

Like the device in chapter 6, this device (demon) is designed as a rigid ring

with inward pointing blades. However, in this model, for symmetry, we attach two blue blades at opposite positions on the ring instead of using only one blade as in chapter 6 (compare FIG 7.1 and FIG 6.1). The device rotates freely around the central axis. We denote the state of the device by θ_D, the angular position of one blade. The blades of the device are frequently hit by the gas particles, and as a result, the device undergoes random rotation. The blades of the device can also collide with the paddle (bit) that is in the vertical range of the gap, which occurs when $\theta_D = \theta_B + N\pi$, for any integer N. We place the device at the vertical range of the gaps and design their size in a way that at any time, there is one and only one paddle that is within the vertical range of the gap, and at the same time the paddle is at the vertical collision range with the blades of the device. Such a paddle is called the *interacting bit*, and its corresponding gate is called the *engaging gate*. Each bit (and gate) spend the same amount of time (interaction interval τ^{bit}) as an interacting bit (and engaging gate). Here, in this model, instead of describing an incoming tape by its fraction of 0's and 1's, we compare the bits with the gates from the preprogrammed reference sequence. It is convenient to characterize the incoming memory sequence by the fraction of the bits that are in agreement and in disagreement with their gates. Here we define the excess ratio of agreement as

$$\delta = P_{\text{in}}(\text{same}) - \text{P}_{\text{in}}(\text{diff}). \tag{7.1}$$

We claim that, if all of the incoming bits agree with the preprogrammed gates ($\delta = 1$), then the device favors CCW rotation. By symmetry, the device favors CW

rotation if $\delta = -1$.

External loads of various magnitude can be attached to the device. We denote the load by torque Γ, which is positive when it favors CW rotation.

7.3 Modes of the device

Assume the demon is programmed with a particular sequence of gates such as the binary digits of π, and assume an incoming sequence of bits are prepared in accordance with the corresponding sequence of gates. If the sequence of bits is perfectly prepared without randomness or uncertainty, then each bit is in agreement with its corresponding gate. More generally, we allow for uncorrelated uncertainties in the incoming sequence of bits, which are described by $P_{\text{in}}(\text{same})$ and $P_{\text{in}}(\text{diff})$ – the probabilities of agreement and disagreement between the bit and its corresponding gate. Depending on $\delta = P_{\text{in}}(\text{same}) - \mathrm{P}_{\text{in}}(\text{diff})$, the temperature of the heat bath T, and the magnitude of the external load Γ, this device can function in three different modes.

As in Eq. 6.1, we can define the Shannon entropy of an incoming bit as

$$S_{\text{in}} = -\big[P_{\text{in}}(\text{same}) \ln \mathrm{P}_{\text{in}}(\text{same}) + \mathrm{P}_{\text{in}}(\text{diff}) \ln \mathrm{P}_{\text{in}}(\text{diff})\big] \tag{7.2}$$

and the Shannon entropy of an outgoing bit as

$$S_{\text{out}} = -\big[P_{\text{out}}(\text{same}) \ln \mathrm{P}_{\text{out}}(\text{same}) + \mathrm{P}_{\text{out}}(\text{diff}) \ln \mathrm{P}_{\text{out}}(\text{diff})\big] \tag{7.3}$$

7.3.1 The engine mode

Like the mechanical demon described in Sec. 6.2.1, the device is able to work as an information engine – it converts heat from the heat bath into work against an external load, while writing information onto a memory sequence. Note that the process of writing information onto a clean sequence appears differently in this programmable demon than in the earlier mechanical demon. In the earlier model, the memory sequence that moves past the demon changes from a pure sequence of 0's into a mixture of 0's and 1's. Here as a programmable demon writes information onto a clean sequence, both the incoming and the outgoing sequences can consist of a mixture of 0's and 1's. The only difference between the incoming and outgoing sequence is that compared with the incoming bits, the outgoing bits have greater uncertainty, i.e. less correlation with the preprogrammed gates. In other words, the uncertainty (or entropy) of the bits is increased by the demon as it writes information.

Assuming that the incoming sequence is perfectly prepared in accordance with the programmed gates, $P_{\text{in}}(\text{same}) = 1$, we can show that the device favors CCW rotation, which can be harnessed to perform work against a small external load. In this case, there are only two possible combinations of an incoming bit and its corresponding gate – $(0, \bar{0})$ and $(1, \bar{1})$. These two cases are shown in the top view of the system sketched in Fig. 7.2. In the first case, an incoming bit initially confined in $\theta_B \in [0, \pi)$ tends to spontaneously expand to the full circle $[0, 2\pi)$ through the gaps in the red bar. But the expansion via the gap at $\theta = 0$ is blocked by the rigid

gate $\bar{0}$. As a result, the expansion of the bit can only pass through the gap at $\theta = \pi$, thereby favoring CCW rotation. The expanding paddle then carries the device to rotate CCW due to the collisions between the paddle and the blades. Similarly, if an incoming bit is initially in state 1, then its expansion favors CCW rotation as the gate $\bar{1}$ blocks the gap at $\theta = \pi$. A small external load can be added against the CCW rotation to extract work from this rectified thermal fluctuation. Note that the expansions of the paddles are isothermal, since they are immersed in the heat bath. In other words, the energies of the paddles do not change on average, and the energy used to perform mechanical work comes solely from the heat bath.

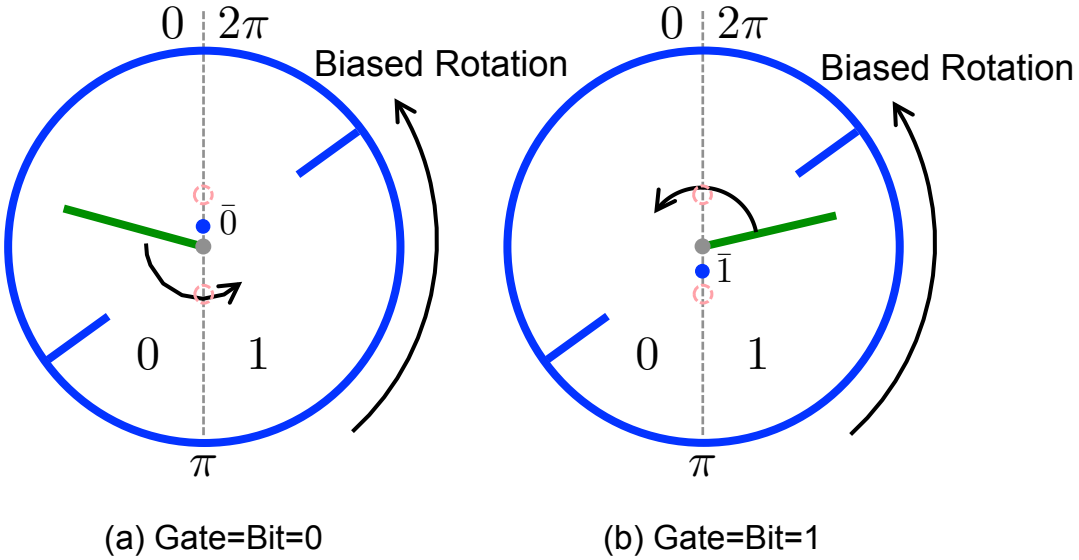


Figure 7.2: Engine Mode. The demon prefers CCW rotation when the bit starts with the state that is in agreement with its corresponding gate. The blue dots represent the programmed gates.

7.3.2 The eraser mode

Like the earlier mechanical demon (see Sec. 6.2.2), the programmable demon can work as a Landauer's eraser. However, in contrast with the earlier mechanical

demon, where a random incoming sequence is erased to a pure sequence of 0's, or a pure sequence of 1's, this device can erase a random incoming sequence into any sequence, depending on the programmed reference sequence. In other words, this eraser can be considered as a copier, which takes in any random incoming sequence and produce a truthful copy of the preprogrammed reference sequence. This erasing process decreases the Shannon entropy of the sequence of bits, which is compensated by work dissipated into the heat bath, which increases the entropy of the heat bath.

This mode can be achieved by providing a random incoming sequence to the demon and attaching a very heavy external load. As sketched in Fig. 7.3, the heavy external load is able to force any incoming bit into the state of its corresponding gate, that is, gate= $\bar{0}$ in Fig. 7.3 (a) and gate= $\bar{1}$ in Fig. 7.3 (b).

In Fig. 7.3 (a), a very heavy external load is attached to the device, and it strongly favors CW rotation. Any bit (paddle) at the interaction interval, regardless of its state (0 or 1), is forced by the CW motion of the device, and is finally pinched between a blade of the device and the rigid gate. After one period τ^{bit}, this paddle leaves the device, in state 0. Similarly, if the programmed gate is in state $\bar{1}$, shown as Fig. 7.3 (b), then the paddle is pinched at state 1 until it leaves the device and becomes an outgoing bit. Over time, the device driven by the strong external load will perform CW rotations, and work is continuously performed by the load and dissipated into the heat bath. As a result, the randomness of incoming sequence is removed while energy is dissipated into a heat bath. From an engineering point of view, this eraser can be used to as a sequence copier that prepare any random sequence into a determined target sequence.

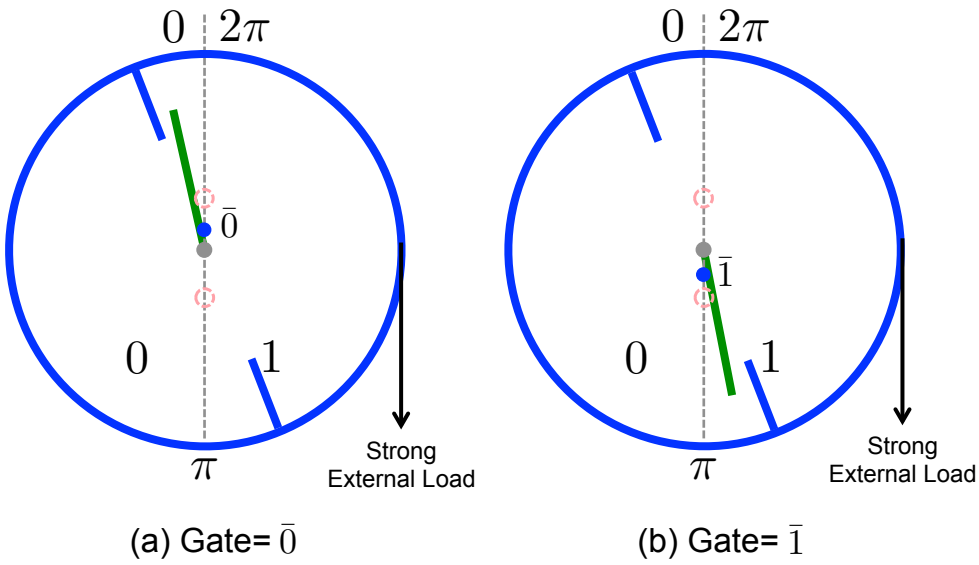


Figure 7.3: Eraser Mode. Under strong external load, the bits are pinched at the states that are in agreement with the states of the gates.

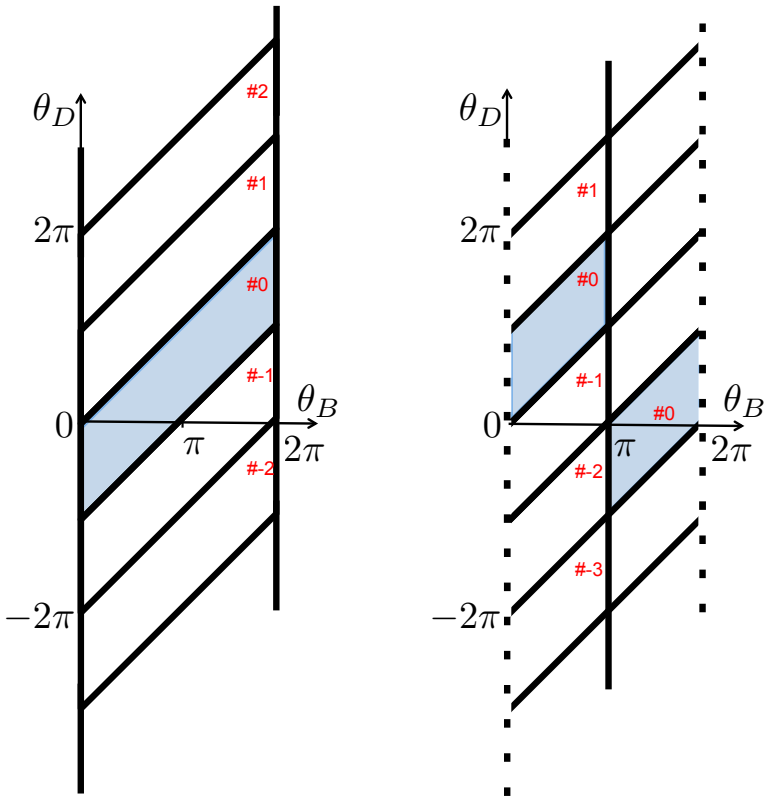
7.3.3 The dud mode

Similar to Sec. 6.2.3, the programmable Maxwell's demon has a mode that is useless. In such a mode, the device dissipates work into the surrounding bath, while the uncertainty contained in the memory sequences increases.

7.4 Quantitative analysis of the programmable Maxwell's demon

In chapter 6, we designed a mechanical Maxwell's demon and studied the dynamics of the composite system of the demon and its interacting bit. Additionally, we showed that such dynamics can be considered as a random walk along a one dimensional lattice of cells. (Refer to Sec. 6.3.) Here we use the same strategy to analyze the programmable Maxwell's demon.

In this section, we solve for the average work done by the device, and the



(a) Gate $= \bar{0}$ (b) Gate $= \bar{1}$

Figure 7.4: The configuration space of the composite state of the demon and the interacting bit. The bold solid lines represent hard wall boundaries and the bold dashed lines represent periodic boundary conditions. The configuration space for a system with gate at $\bar{0}$ is sketched on the left (a). The configuration space with a gate at $\bar{1}$ is on the right (b). At the end of an interaction interval, a new bit replaces the old bit and a new engaging gate replaces the old one. When the new engaging gate is in a different state from the old gate, at the end of the interaction interval, the change of the gate causes a instantaneous change of the boundary conditions from (a) to (b) or (b) to (a). For convenience, we index the lattice of cells by their range of θ_D. Note that at the same index, the cell shown in (a) (for gate $\bar{0}$) is equivalent to the cell shown in (b) (for gate $\bar{1}$), which can be easily verified by considering the periodic boundary condition.

Shannon entropy change of the memory tape per bit. The state of the composite system consisting of the demon and the interacting bit is denoted by (θ_B, θ_D). The configuration space and boundary conditions of the composite system are sketched in Fig. 7.4. Notice that the boundary conditions now depend on the state of the engaging gate.

During each interaction interval τ^{bit}, the programmable demon and the interacting bits are able to evolve randomly but are confined within a cell, whose boundary is defined by the hard collisions between the gate and the bit, and between the demon and the bit. From an ensemble point of view, if the composite system starts at initial states within the same cell #0, the random evolution that occurs during a single interaction interval relaxes the composite system into a Boltzmann distribution restricted by cell #0. At the end of the interaction interval, both the bit and its corresponding gate are suddenly replaced by a new bit and a new gate. As in Sec. 6.3, the bit renewal maps the distribution of the composite system into a different distribution depending on the state of the new bit. In addition to the bit renewal, the state of the gate may also change instantaneously. Such a *gate renewal* instantaneously changes the boundary conditions of the composite system. The probability distributions of the composite system, and the boundary conditions right after bit renewal and gate renewal are shown in Fig. 7.5.

With the above discussion, one can treat the dynamics of θ_D as a discrete time random walk along a discretized lattice of cells. If we assume the gate and the bit are in the same state, at each time step, the composite system can jump up by 1 or 2 cells, which on average corresponds to a CCW rotation of a half circle or a full

circle; it can also jump down by one cell, which corresponds to a CW rotation of a half circle.

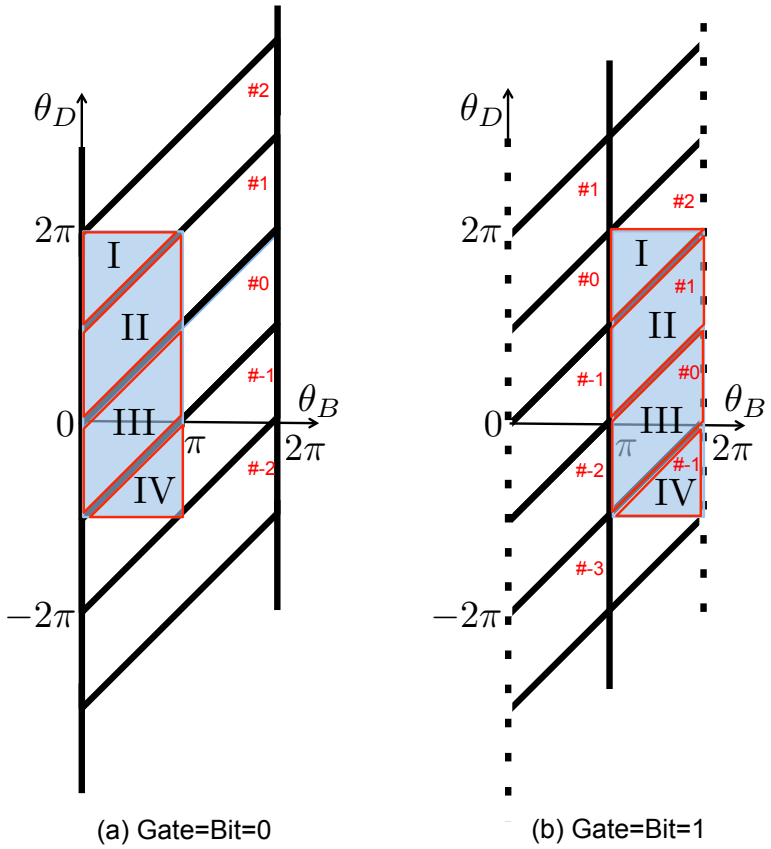


Figure 7.5: The shaded regions indicate the support of the distribution of a composite system right after bit renewal and gate renewal. In (a), the new bit is in state 0, and the gate is in state $\bar{1}$. In (b), the new bit is in state 1, and the gate is in state $\bar{1}$. We have assumed that the composite system is confined in cell #0 right before the bit renewal. In both (a) and (*b*) we observe that the composite system can jump up by -1,0,1, or 2 cells from its original cell #0.

As with the model in Chapter 6, this problem is analytically solvable when we assume $\tau^{\text{bit}} \gg \tau^{\text{eq}}$, so that the composite system equilibrates with the heat bath before each bit renewal. Using a strategy similar to that in Sec. 6.3, we can compute the probability of the jumps conditioned on the agreement or disagreement between the state of the bit and its corresponding gate. At the bit renewal, we coarse grain the dynamics of the composite system into the following events: (1) state stays in

the same cell; (2) state jumps up by 1 cell; (3) state jumps down by 1 cell; (4) state jumps up by 2 cells; (5) state jumps down by 2 cells. Depending on the incoming bit, the incoming gate, the strength of the external load and the temperature of the heat bath, we can compute the probability of these different events. We summarize our results in the following.

When the incoming bit agrees with its gate (e.g. bit 1 and gate $\bar{1}$ or bit 0 and gate $\bar{0}$), we have

$$P_{+2}^{\text{same}} \equiv P_{+2}^{0,\bar{0}} = P_{+2}^{1,\bar{1}} = \frac{\pi\beta\Gamma + e^{\pi\beta\Gamma}(\pi\beta\Gamma - 2) + 2}{\pi\beta\Gamma\left(e^{\pi\beta\Gamma} - 1\right)^2\left(e^{\pi\beta\Gamma} + 1\right)} \tag{7.4}$$

$$P_{+1}^{\text{same}} \equiv P_{+1}^{0,\bar{0}} = P_{+1}^{1,\bar{1}} = \frac{e^{\pi\beta\Gamma}\left[\pi\beta\Gamma\left(e^{\pi\beta\Gamma} - 3\right) + 2\right] - 2}{\pi\beta\Gamma\left(e^{\pi\beta\Gamma} - 1\right)^2\left(e^{\pi\beta\Gamma} + 1\right)} \tag{7.5}$$

$$P_{\text{stay}}^{\text{same}} \equiv P_{\text{stay}}^{0,\bar{0}} = P_{\text{stay}}^{1,\bar{1}} = \frac{e^{\pi\beta\Gamma}\left[\pi\beta\Gamma + e^{\pi\beta\Gamma}\left(-3\pi\beta\Gamma + 2e^{\pi\beta\Gamma} - 2\right)\right]}{\pi\beta\Gamma\left(e^{\pi\beta\Gamma} - 1\right)^2\left(e^{\pi\beta\Gamma} + 1\right)} \tag{7.6}$$

$$P_{-1}^{\text{same}} \equiv P_{-1}^{0,\bar{0}} = P_{-1}^{1,\bar{1}} = \frac{e^{2\pi\beta\Gamma}\left[\pi\beta\Gamma + e^{\pi\beta\Gamma}(\pi\beta\Gamma - 2) + 2\right]}{\pi\beta\Gamma\left(e^{\pi\beta\Gamma} - 1\right)^2\left(e^{\pi\beta\Gamma} + 1\right)} \tag{7.7}$$

and

$$P_{-2}^{\text{same}} \equiv P_{-2}^{0,\bar{0}} = P_{-2}^{1,\bar{1}} = 0 \tag{7.8}$$

Alternatively, if the next incoming bit is not in agreement with its gate, we can find

$$P_{+2}^{\text{diff}} \equiv P_{+2}^{0,\bar{1}} = P_{+2}^{1,\bar{0}} = 0 \tag{7.9}$$

$$P_{+1}^{\text{diff}} \equiv P_{+1}^{0,\bar{1}} = P_{+1}^{1,\bar{0}} = \frac{\pi\beta\Gamma + e^{\pi\beta\Gamma}(\pi\beta\Gamma - 2) + 2}{\pi\beta\Gamma\left(e^{\pi\beta\Gamma} - 1\right)^2\left(e^{\pi\beta\Gamma} + 1\right)} \tag{7.10}$$

$$P_{\text{stay}}^{\text{diff}} \equiv P_{\text{stay}}^{0,\bar{1}} = P_{\text{stay}}^{1,\bar{0}} = \frac{e^{\pi\beta\Gamma}\left[\pi\beta\Gamma\left(e^{\pi\beta\Gamma} - 3\right) + 2\right] - 2}{\pi\beta\Gamma\left(e^{\pi\beta\Gamma} - 1\right)^2\left(e^{\pi\beta\Gamma} + 1\right)} \tag{7.11}$$

$$P_{-1}^{\text{diff}} \equiv P_{-1}^{0,\bar{1}} = P_{-1}^{1,\bar{0}} = \frac{e^{\pi\beta\Gamma}\left[\pi\beta\Gamma + e^{\pi\beta\Gamma}\left(-3\pi\beta\Gamma + 2e^{\pi\beta\Gamma} - 2\right)\right]}{\pi\beta\Gamma\left(e^{\pi\beta\Gamma} - 1\right)^2\left(e^{\pi\beta\Gamma} + 1\right)} \tag{7.12}$$

and

$$P_{-2}^{\text{diff}} \equiv P_{-2}^{0,\bar{1}} = P_{-2}^{1,\bar{0}} = \frac{e^{2\pi\beta\Gamma}\left[\pi\beta\Gamma + e^{\pi\beta\Gamma}(\pi\beta\Gamma - 2) + 2\right]}{\pi\beta\Gamma\left(e^{\pi\beta\Gamma} - 1\right)^2\left(e^{\pi\beta\Gamma} + 1\right)} \tag{7.13}$$

Observing the demon for a long time (or many periods), the evolution of the composite system can be modeled as a discrete time random walk with time step τ^{bit}. During each period τ^{bit}, the composite system can jump by -2,-1,0,+1,+2 cells, corresponding to a rotation of $-2\pi,\ -\pi,\ 0,\ \pi,\ 2\pi$. Considering the ratio of bit-gate agreement in the incoming tape, we can compute the average probability of jumps of the random walk.

$$R_{-2} = P_{\text{in}}(\text{same}) \cdot P_{-2}^{\text{same}} + P_{\text{in}}(\text{diff}) \cdot P_{-2}^{\text{diff}} \tag{7.14}$$

$$R_{-1} = P_{\text{in}}(\text{same}) \cdot P_{-1}^{\text{same}} + P_{\text{in}}(\text{diff}) \cdot P_{-1}^{\text{diff}} \tag{7.15}$$

$$R_{+1} = P_{\text{in}}(\text{same}) \cdot P_{+1}^{\text{same}} + P_{\text{in}}(\text{diff}) \cdot P_{+1}^{\text{diff}} \tag{7.16}$$

and

$$R_{+2} = P_{\text{in}}(\text{same}) \cdot P_{+2}^{\text{same}} + P_{\text{in}}(\text{diff}) \cdot P_{+2}^{\text{diff}} \tag{7.17}$$

Over many interaction intervals, the random walk may be biased to perform a systematic rotation, which can be used to perform work against external load. The average work done during each period is simply the average rotation of the demon per period multiplied by the external torque:

$$\begin{aligned} W &= 2\pi\Gamma \cdot R_{+2} + \pi\Gamma \cdot R_{+1} - \pi\Gamma \cdot R_{-1} - 2\pi\Gamma \cdot R_{-2} \\ &= \frac{\pi\beta\Gamma\delta - \pi\beta\Gamma\left[3\coth(\pi\beta\Gamma) + \operatorname{csch}(\pi\beta\Gamma)\right] + 4}{2\beta} \end{aligned} \tag{7.18}$$

As a consistency check, let us consider the following limits. When the external load is extremely small ($\Gamma \to 0$), then to the leading order in Γ we get $W = \delta\pi\Gamma/2$ as in Sec. 6.3. As expected, no work is done in the absence of an external load. When the load is very large ($\Gamma \gg k_B T$), then we expect that in each interaction interval the demon completes either a half circle of CW rotation (if the incoming bit agrees with its gate) or a full circle of CW rotation (if the incoming bit disagrees with its gate). In agreement with this expectation, in the limit $\Gamma \to +\infty$, Eq. 7.18 asymptotically becomes $W = (\delta - 3)\pi\Gamma/2$. See similar comments following Eq. 6.16.

Additionally, we can compute the fractions of bit-gate agreement and disagreement in the outgoing tape:

$$P_{\text{out}}(\text{same}) = 1 - \frac{1}{e^{\pi\beta\Gamma} + 1} \quad , \tag{7.19}$$

$$P_{\text{out}}(\text{diff}) = \frac{1}{\mathrm{e}^{\pi\beta\Gamma}+1} \quad . \tag{7.20}$$

In the absence of an external load, the outgoing bits are uncorrelated with the reference sequence. When the load is large ($|\Gamma| \gg k_B T$), the outgoing bits are either perfectly correlated or perfectly anti-correlated with the reference sequence, depending on the sign of Γ. See similar comments following Eq. 6.18. The Shannon entropy change of the memory tape per period is

$$\begin{aligned}\Delta S_{\text{bit}} = &-\frac{\left(e^{\pi\beta\Gamma}+1\right)\left[(\delta-1)\log(1-\delta)-(\delta+1)\log(\delta+1)+\log(4)\right]}{2\left(e^{\pi\beta\Gamma}+1\right)} \\ &-\frac{e^{\pi\beta\Gamma}\log\left(1-\frac{1}{e^{\pi\beta\Gamma}+1}\right)-\log\left(e^{\pi\beta\Gamma}+1\right)}{\left(e^{\pi\beta\Gamma}+1\right)}\end{aligned} \tag{7.21}$$

By tuning the combination of β, Γ and δ, we can change the behavior of the device among the three modes. The phase diagram of the demon can be obtained by solving for the zeros of the average work, and entropy change (see eqs. 7.18 and 7.21). The phase diagram is shown in Fig. 7.6.

As in Section 6.4, we can define the efficiencies of the engine mode and the eraser mode. When the device functions as an eraser,

$$W < k_B T \Delta S_{\text{bit}} < 0 \tag{7.22}$$

and the efficiency is defined as

$$\eta_{\text{eraser}} = \frac{k_B T \Delta S_{\text{bit}}}{W} \leq 1 \tag{7.23}$$

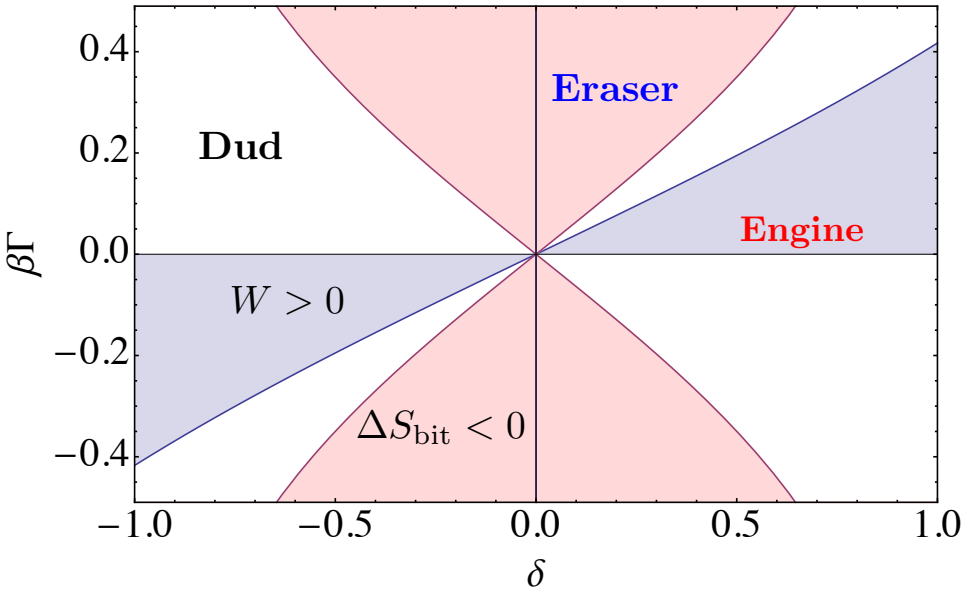


Figure 7.6: The phase diagram of the programmable Maxwell's demon.

When the device functions as an engine,

$$k_B T \Delta S_{\rm bit} > W > 0 \tag{7.24}$$

and the efficiency is defined as

$$\eta_{\rm engine} = \frac{W}{k_B T \Delta S_{\rm bit}} \leq 1 \tag{7.25}$$

When the device functions in the dud mode,

$$W < 0 < k_B T \Delta S_{\rm bit}. \tag{7.26}$$

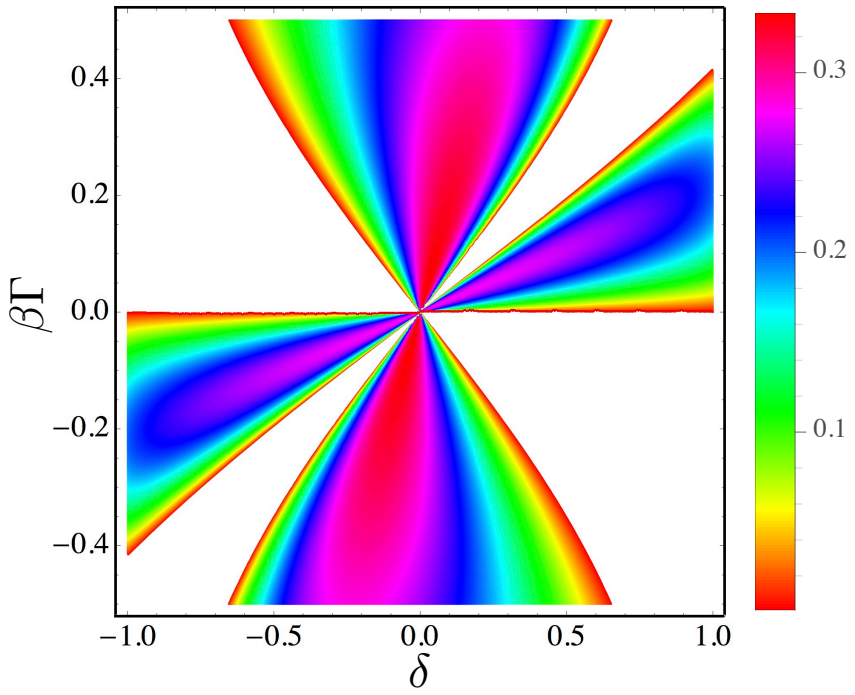


Figure 7.7: Efficiency plot of the programmable demon, setting $k_B = 1$ and $\tau^{\text{bit}} \gg \tau^{\text{eq}}$.

7.5 Discussion

7.5.1 Comparison between two models of Maxwell's demon

In this section, we compare the programmable demon with the mechanical demon (as described in Chapter 6). To simplify the comparison, let us prepare the programmable demon by setting every preprogrammed gate to $\bar{0}$. In this case, the quantity δ defined in Eq. 7.1 becomes the excess ratio of 0's in the incoming sequence of bits, which is the same as the quantity δ defined for the mechanical demon in Eq. 6.2. When the temperature of the heat bath and the external loads are prepared identically, the two models are almost identical, except that the numbers of blade(s) attached to the demons are different. If the two models behaves differently, it is

caused by the difference in the number of blades.

In the limit that $\tau_{\text{bit}} \gg \tau_{\text{eq}}$, at the end of each interaction interval, the marginal distribution of θ_B is only a function of temperature and the external load, and it is independent of the number of blades. As a result, the ratio of 0's in the outgoing sequence is also independent of the number of blades (see Eqs. 6.17 and 7.19). In other words, the Shannon entropy production (or reduction) of the sequence is the same in both models (see Eqs. 6.19 and 7.21). We can compare Fig. 6.7 and Fig. 7.6 and verify that the domains of the eraser mode ($\Delta S_{\text{bit}} < 0$) for both models are identical.

In either model, the demon interacts with the interacting bit through the collisions between the blade(s) and the paddle. Such interactions are slack: On the mechanical demon, there is only one blade, thus even if the paddle is fixed at a certain angle, the demon can still rotate freely around a full circle (2π). We call this range of motion the *slack* between the demon and the interacting paddle. On the programmable demon, there are two blades, and the slack between the demon and the paddle is π. It turns out that with less slack, the programmable demon delivers more work than the mechanical demon when operating under the same conditions. Comparing the phase diagrams, Fig. 6.8 and Fig. 7.7, we find that the domain of the engine mode of the programmable demon is larger than that of the mechanical demon. In addition, we find that the efficiency of the programmable demon is generally higher than that of the mechanical demon.

7.5.2 Demon with infinitely many blades

Inspired by the discussion of slack and the ability to deliver work, it is natural to design and study a model where the slack between the demon and the bit reduces to 0, which is expected to deliver more work. This can be done by increasing the number of blades to infinity, while keeping the total thickness of the blades to be 0. (We assume that the blades are evenly spaced on the demon, and each blade is infinitely thin.) In this way, the vertical downward motion of the incoming bits (paddles) is not blocked by the blade, and the rotation of the demon paces perfectly with the rotation of the interacting paddle. In this case, we can carry out calculations similar to those described in Secs. 6.3 and 7.4 to compute the average work performed by the demon per bit:

$$W = \frac{1}{\beta} + \frac{\pi\Gamma\delta}{2} - \pi\Gamma \coth(\pi\beta\Gamma) \tag{7.27}$$

As a consistency check, when the external load is absent ($\Gamma \to 0$), the work per bit is equal to 0. When the external load is very large ($\Gamma \gg k_B T$), the work done per bit becomes $(\delta - 2)\pi\Gamma/2$. See similar comments following Eqs. 6.16 and 7.18.

As discussed in Sec. 7.5.1, the Shannon entropy change per bit is independent of the number of blades. Thus the Shannon entropy change per bit in this infinite-blade model is identical with the values of Eq. 6.19 and Eq. 7.21. We plot the efficiency diagram (and phase diagram) of this demon in Fig. 7.8. As expected, this demon has a generally higher efficiency than the two other demons. In addition, it

has a larger domain of the engine mode in the phase diagram.

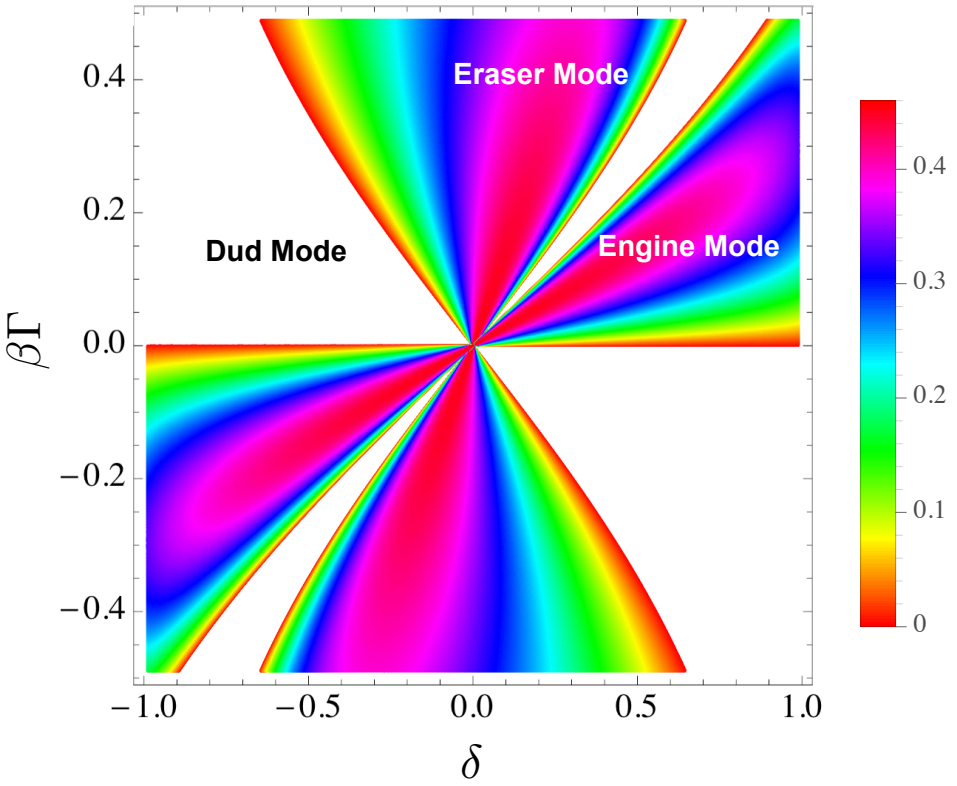


Figure 7.8: Efficiency plot of the programmable demon with infinite number of blades, setting $k_B = 1$ and $\tau^{\text{bit}} \gg \tau^{\text{eq}}$.

Chapter 8 : Summary and Future Outlook

8 .1 Computing free energies of crystalline solids

Direct numerical evaluations of crystalline free energies (or partition functions) suffer from high computational cost in evaluating ND-dimensional integrals over the configuration space. Here N is the number of particles in the crystal and D is the spatial dimensionality. Typically, numerical methods are developed to evaluate the free energy difference between a system of interest and a reference system. In this thesis work, I described a new method that numerically evaluates the ND-dimensional integral by introducing a convenient order parameter.

The new method is not expected to be significantly more efficient than the Frenkel-Ladd (FL) method or the lattice switch (LS) method. For both the new method and the two older methods, it is crucial to obtain adequate sampling over sets of configurations that are closely confined near the perfect (reference) lattice configuration. We call these sets of configurations the *reference region*. In the FL method, the reference region consists of the typical configurations of the Einstein crystal (with $\lambda = \lambda_E$ in Eq. 1.17). In the LS method, the reference region is the gateway region where the lattice switch operations take place. In the method described in this thesis, the reference region is $\mathbb{S}(\bar{r})$ – the hypersphere (of a small radius $\bar{r}$)

around the perfect lattice configuration. It is where $P_0(\bar{r})$ is evaluated. These reference regions share the similarity that they are rarely visited when directly sampling the crystalline solids. To reduce the computational cost, advanced sampling schemes are used in these three methods for adequately sampling the reference regions, and their efficiencies are not expected to differ significantly.

A natural next step of research is trying to develop numerical methods that are significantly more efficient than FL, LS and the method described in this thesis. To achieve this, one should get around the sampling difficulty caused by the reference region. This can possibly be achieved by changing the reference region so that it is more frequently sampled in a direct MC simulation. A preliminary attempt is made in Sec. 2.4.2, but it is far from adequate. It is still an open question to develop methods that significantly increase the numerical efficiency.

8 .2 Engineering Maxwell's demons

In this thesis, I have described two models of autonomous Maxwell's demon. Both models describe mechanical devices that rectify the thermal energy from a single heat bath to lift a mass, while writing information on a memory register.

A future engineering challenge is to expand the design of the memory sequences used in the information engine. In the models described in this thesis, the memory registers are modeled by sequences of binary digits (0's and 1's). In principle, the digits in a memory register can be more than binary. For example, one can consider a memory register consists of the 26 characters of the English alphabet.

It is particularly interesting to try to engineer an information engine that rectifies thermal motion to lift a mass while writing a sequence of English characters. In other words, can one design a programmable demon that performs maximum work when it is provided a sequence of characters taken from Shakespeare's Sonnets [65]? To achieve this goal, one can increase the number of red blocking bars (shown as red dots in Fig. 8 .1) so that the angular position of each paddle is separated into multiple states. Then, the next step is to engineer the demon and the preprogrammed gates accordingly so that the device operates as expected. If such a machine is designed, it can be called an *information turbine*.

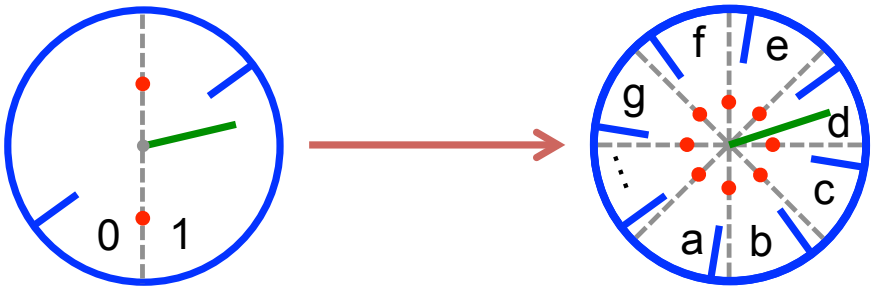


Figure 8 .1: Upgrading memory from binary states (0 and 1) to multiple states (a, b, c...). The demon on the left operates on avsequence of binary bits; the demon on the right is expected to operate on avsequence of English characters. Due to the limited space, not all of the 26 characters are shown here.

In addition to upgrading the memory register, it is also interesting to find models of autonomous Maxwell's demons that operate under the maximum efficiency that is allowed by the second law. In this case, the work performed by the demon is equal to the change of the Shannon entropy in the memory multiplied by $k_B T$. To achieve the maximum efficiency, the model must remain in equilibrium with the heat bath at any instant in time. This is left as an open challenge for future engineers.

Appendix A: Measure of the central island $\mathcal{C}$

To derive Eqs. 2.23-2.26, it will be useful to define a $(3N-3)$-dimensional hyperplane,

$$\mathbb{L} \equiv \left\{ \mathbf{r} \in \mathbb{R}^{3N} \;\middle|\; \sum_{i=1}^{N} \vec{r}_i = \vec{0} \right\}. \tag{A1}$$

Conformation space is a bounded region $\mathbb{K} \subset \mathbb{L}$ specified by the conditions $\vec{r}_i \in \Omega$, $i = 2, \cdots, N$ (see Chapter 2). The central island $\mathcal{C}$ is in turn a region within conformation space:

$$\mathcal{C} \subset \mathbb{K} \subset \mathbb{L} \tag{A2}$$

For any conformation $\mathbf{r} \in \mathbb{K}$, we have defined an indicator function $\Theta_{\mathcal{C}}(\mathbf{r})$ whose value is 1 or 0 according to whether or not $\mathbf{r}$ belongs to the central island. We now extend the domain of this indicator function to include the entire hyperplane $\mathbb{L}$, by setting $\Theta_{\mathcal{C}}(\mathbf{r}) = 0$ for all $\mathbf{r} \in \mathbb{L}$ that fall outside of $\mathbb{K}$. Then for a hard-sphere crystal, Eq. 2.22 becomes

$$Z_{\mathrm{FCC}} = \frac{|\Omega|}{\Lambda^{3N}} \int_{\mathbb{R}^{3N}} \mathrm{d}\mathbf{r}\, \Theta_{\mathcal{C}}(\mathbf{r})\, \delta^3 \left(\sum_{i=1}^{N} \vec{r}_i \right) \tag{A3}$$

as $e^{-\beta V} = 1$ for any $\mathbf{r} \in \mathcal{C}$. We can write this more compactly as

$$Z_{\text{FCC}} = \frac{|\Omega|}{\Lambda^{3N}} \int_{\mathbb{L}} \mathrm{d}\mu \, \Theta_{\mathcal{C}} \tag{A4}$$

since integration with the δ^3-function in Eq. A3 induces a measure μ on the hyperplane $\mathbb{L}$. The integral appearing here is the total measure (or "volume") of the central island $\mathcal{C}$.

Using the order parameter $r \equiv |\mathbf{r}|$ to denote the radius, let us define a continuous family of concentric hyperspheres within the hyperplane $\mathbb{L}$,

$$\mathbb{S}(r) \equiv \left\{ \mathbf{r} \in \mathbb{L} \,\middle|\, |\mathbf{r}| = r \right\} \quad , \quad 0 \leq r < \infty \tag{A5}$$

Now we rewrite the partition function (Eq. A3) as a radial integral:

$$\begin{aligned} Z_{\text{FCC}} &= \frac{|\Omega|}{\Lambda^{3N}} \int_0^\infty \mathrm{d}r \int_{\mathbb{R}^{3N}} \mathrm{d}\mathbf{r} \, \Theta_{\mathcal{C}}(\mathbf{r}) \, \delta^3 \left(\sum_{i=1}^N \vec{r}_i \right) \, \delta \left(r - |\mathbf{r}| \right) \\ &= \frac{|\Omega|}{\Lambda^{3N}} \int_0^\infty \mathrm{d}r \, P_0(r) s(r) \end{aligned} \tag{A6}$$

where

$$P_0(r) = \frac{\int_{\mathbb{R}^{3N}} \mathrm{d}\mathbf{r} \, \Theta_{\mathcal{C}}(\mathbf{r}) \, \delta^3 \left(\sum_{i=1}^N \vec{r}_i \right) \, \delta(r - |\mathbf{r}|)}{\int_{\mathbb{R}^{3N}} \mathrm{d}\mathbf{r} \, \delta^3 \left(\sum_{i=1}^N \vec{r}_i \right) \, \delta(r - |\mathbf{r}|)} = \frac{\int_{\mathbb{L}} \mathrm{d}\mu \; \Theta_{\mathcal{C}} \, \delta(r - |\mathbf{r}|)}{\int_{\mathbb{L}} \mathrm{d}\mu \; \delta(r - |\mathbf{r}|)} \tag{A7}$$

$$s(r) = \int_{\mathbb{R}^{3N}} \mathrm{d}\mathbf{r} \, \delta^3 \left(\sum_{i=1}^N \vec{r}_i \right) \, \delta(r - |\mathbf{r}|) = \int_{\mathbb{L}} \mathrm{d}\mu \, \delta(r - |\mathbf{r}|) \tag{A8}$$

Eqs. A6 - A8 are equivalent to Eqs. 2.23-2.26 of the main text. The quantity $P_0(r)$

is equal to the average value of $\Theta_{\mathcal{C}}(\mathbf{r})$ over the hypersphere $\mathbb{S}(r)$ – i.e. it is the probability to land on a point within $\mathcal{C}$, when sampling uniformly from $\mathbb{S}(r)$ (see Fig. 2.3) – and $s(r)$ is the total measure of this hypersphere. We can loosely think of $s(r)$ as the "surface area" of the hypersphere, keeping in mind that this area is defined with respect to the measure μ.

To evaluate $s(r)$ we will perform a rotation of coordinate axes. First, we write the vector $\mathbf{r} = (r_{1x}, \cdots, r_{Nz})$ as a linear combination of "old" unit basis vectors that point along the x, y, and z coordinate axes, namely

$$\mathbf{r} = r_{1x} \cdot \hat{x}_1 + \cdots + r_{Nz} \cdot \hat{z}_N \tag{A9}$$

The same vector is expressed using a "new" orthonormal basis set, $\{\hat{u}_i\}$:

$$\mathbf{r} = u_1 \cdot \hat{u}_1 + \cdots + u_{3N} \cdot \hat{u}_{3N} \tag{A10}$$

The mapping from the old to the new basis set is a rotation in $3N$-space. We take the first three vectors of the new basis set to be:

$$\hat{u}_1 = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \hat{x}_i \quad , \quad \hat{u}_2 = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \hat{y}_i \quad , \quad \hat{u}_3 = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \hat{z}_i \tag{A11}$$

(Note that these are normalized to unity, and mutually orthogonal.) We will not need explicit expressions for the remaining basis vectors $\hat{u}_4, \cdots \hat{u}_{3N}$. The first three

coefficients u_1, u_2 and u_3 in Eq. A10 are:

$$u_1 = \mathbf{r} \cdot \hat{u}_1 = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} r_{ix} \quad , \quad u_2 = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} r_{iy} \quad , \quad u_3 = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} r_{iz} \tag{A12}$$

This allows us to perform the integral over the δ^3-function in Eq. A8:

$$\begin{aligned} s(r) &= \int \mathrm{d}\mathbf{r} \;\; \delta(r - |\mathbf{r}|) \;\; \delta^3 \left(\sum_{i=1}^{N} \vec{r}_i \right) \\ &= \int \mathrm{d}\mathbf{u} \;\; \delta(r - |\mathbf{u}|) \;\; \delta(\sqrt{N} u_1) \;\; \delta(\sqrt{N} u_2) \;\; \delta(\sqrt{N} u_3) \\ &= \frac{1}{N^{3/2}} \int \mathrm{d}u_4 \cdots \mathrm{d}u_{3N} \;\; \delta \left(r - \sqrt{u_4^2 + \cdots + u_{3N}^2} \right) \\ &= \frac{1}{N^{3/2}} \, \frac{n \pi^{n/2}}{\Gamma(n/2 + 1)} \, r^{n-1} \quad . \end{aligned} \tag{A13}$$

Here $n = 3N - 3$, $\Gamma(\cdot)$ is the Gamma function, and the final line follows from the Cartesian surface area of a hypersphere of radius r in n dimensions.

Bibliography

[1] S. Pronk and D. Frenkel, J. Chem. Phys. **110**(9), 4589 (1999).

[2] Wikipedia, *Brownian ratchet — wikipedia, the free encyclopedia* (2015), [Online; accessed 26-October-2015], `https://en.wikipedia.org/w/index.php?title=Brownian_ratchet&oldid=682873704`.

[3] D. Mandal and C. Jarzynski, Proc. Natl. Acad. Sci. U.S.A. **109**(29), 11641 (2012).

[4] B. H. Callen, *Thermodynamics & an Intro. to Thermostatistics* (John wiley & sons, 2006).

[5] J. W. Gibbs, *Elementary principles in statistical mechanics* (Courier Corporation, 2014).

[6] R. K. Pathria and P. D. Beal, *Statistical Mechanics* (Butterworth Heinemann, 1996), second edition ed.

[7] D. Frenkel and A. J. Ladd, J. Chem. Phys. **81**(7), 3188 (1984).

[8] D. Frenkel and B. Smit, *Understanding Molecular Simulation From Algorithms to Application* (Academic Press, 2002).

[9] A. D. Bruce, N. B. Wilding, and G. J. Ackland, Phys. Rev. Lett. **79**(16), 3002 (1997).

[10] A. D. Bruce, A. N. Jackson, G. J. Ackland, and N. B. Wilding, Phys. Rev. E **61**(906–919) (2000).

[11] S. C. Mau and D. A. Huse, Phys. Rev. E **59**(4), 4396 (1999).

[12] P. D. Beale, Phys. Rev. E **66**(3), 036132:1 (2002).

[13] C. D. Barnes and D. A. Kofke, Phys. Rev. E **65**(3), 036709:1 (2002).

[14] M. B. Sweatman, Phys. Rev. E **72**(1), 016711:1 (2005).

[15] F. M. Ytreberg and D. M. Zuckerman, J. Chem. Phys. **124**(10), 104105:1 (2006).

[16] N. G. Almarza, J. Chem. Phys. **126**(21), 211103:1 (2007).

[17] C. Vega and E. G. Noya, J. Chem. Phys. **127**(15), 154113:1 (2007).

[18] M. Müller and K. C. Daoulas, J. Chem. Phys. **128**(2), 024903:1 (2008).

[19] T. Schilling and F. Schmid, J. Chem. Phys. **131**(23), 231102:1 (2009).

[20] C. Chen and Y. Xiao, J. Comput. Chem. **31**(7), 1368 (2010).

[21] J. L. Aragones, E. G. Noya, C. Valeriani, and C. Vega, J. Chem. Phys. **139**(3), 034104:1 (2013).

[22] V. Elser, Phys. Rev. E **89**(5), 052404:1 (2014).

[23] S. G. Moustafa, A. J. Schultz, and D. A. Kofke, J. Chem. Phys. **139**(8), 084105:1 (2013).

[24] R. W. Zwanzig, J. Chem. Phys. **22**, 1420 (1954).

[25] C. H. Bennett, J. Comput. Phys. **22**, 245 (1976).

[26] N. Lu and D. A. Kofke, J. Chem. Phys. **114**(17), 7303 (2001).

[27] D. Wu and D. A. Kofke, J. Chem. Phys. **123**(5), 054103 (2005).

[28] D. Wu and D. A. Kofke, J. Chem. Phys. **123**(8), 084109 (2005).

[29] J. G. Kirkwood, J. Chem. Phys. **3**, 300 (1935).

[30] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, J. Chem. Phys. **21**(6), 1087 (1953).

[31] G. Navascués and E. Velasco, J. Chem. Phys. **132**, 134106:1 (2010).

[32] W. G. Hoover and F. H. Ree, J. Chem. Phys. **49**(8), 3609 (1968).

[33] P. Krishna and A. R. Verma, Phys. Status Solidi B **17**(2), 437 (1966).

[34] A. A. Lebedev, Semicond. Sci. Technol. **21**(6), R17 (2006).

[35] S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, and P. A. Kollman, J. Comput. Chem. **13**(8), 1011 (1992).

[36] M. E. Muller, Commum. ACM **2**(4), 19 (1959).

[37] A. Papoulis and S. U. Pillai, *Probability, random variables, and stochastic processes* (Tata McGraw-Hill Education, 2002).

[38] F. W. J. Olver, D. W. Lozier, R. F. Boisvert, and C. W. Clark, eds., *NIST handbook of mathematical functions* (Cambridge University Press, 2010).

[39] H. Touchette, Physics Reports **478**(1–3), 1 (2009).

[40] J. M. Polson, E. Trizac, S. Pronk, and D. Frenkel, J. Chem. Phys. **112**(12), 5339 (2000).

[41] R. Clausius, The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science **12**(77), 81 (1856).

[42] M. Planck, *Treatise on Thermodynamics (3rd edn) english translated by Alexander OGG* (Longmans, Green, London, UK, 1903).

[43] J. Maxwell, *Theory of heat* (Longmans, London, UK, 1871).

[44] B. N. Roy, *Fundamentals of classical and statistical thermodynamics* (John Wiley & Sons, 2002).

[45] G. Lippmann, Rapp. du Congr. Int. d. Phys. Paris **1**(S. 546) (1900).

[46] M. Smoluchowski, Phys. Zeitshur. **13**, 1069 (1912).

[47] R. P. Feynman, *The Feynman Lectures on Physics*, vol. 1 (Addison-Wesley, 1963).

[48] C. Jarzynski and O. Mazonka, Phys. Rev. E **59**(6), 6448 (1999).

[49] P. Eshuis, K. van der Weele, D. Lohse, and D. van der Meer, Phys. Rev. Lett. **104**(24), 248001 (2010).

[50] C. H. Bennett, Int. J. Theor. Phys. **21**(12), 905 (1982).

[51] C. H. Bennett and L. R, Sci. Am. **253**(1), 48 (1985).

[52] R. Landauer, IBM J. Res. Dev. **5**(3), 183 (1961).

[53] O. Penrose, *Foundations of statistical mechanics: a deductive treatment* (Pergamon Press, Oxford, 1970).

[54] H. Quan, Y. Wang, Y. Liu, C. Sun, and F. Nori, Phys. Rev. Lett. **97**(18), 180402 (2006).

[55] T. Sagawa and M. Ueda, Phys. Rev. Lett. **109**(18), 180602 (2012).

[56] P. Strasberg, G. Schaller, T. Brandes, and M. Esposito, Phys. Rev. Lett. **110**(4), 040601 (2013).

[57] J. M. Horowitz, T. Sagawa, and J. M. Parrondo, Phys. Rev. Lett. **111**(1), 010602 (2013).

[58] A. C. Barato and U. Seifert, EPL **101**(6), 60001 (2013).

[59] D. Mandal, H. Quan, and C. Jarzynski, Phys. Rev. Lett. **111**(3), 030602 (2013).

[60] S. Deffner, Phys. Rev. E **88**(6), 062128 (2013).

[61] Z. Lu, D. Mandal, and C. Jarzynski, Phys. Today **67**, 60 (2014).

[62] Z. Lu, *Mechanical maxwells demon in motion (youtube)* (2014), [Online; accessed 23-November-2015], `https://youtu.be/00TyIShzR6o`.

[63] B. Sherwood (2014), private communication, `http://www.glowscript.org/#/user/Bruce_Sherwood/folder/Pub/program/MaxwellDemon`.

[64] Z. Lu, *Programmable mechanical maxwells demon (youtube)* (2015), [Online; accessed 23-November-2015], `https://youtu.be/LkYljJ__-Cs`.

[65] W. Shakespeare, G. B. Evans, and A. Hecht, *The sonnets*, vol. 26 (Cambridge University Press, 1996).